



ANEXO TÉCNICO

DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA  
COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO

Página

SENADO DE LA REPÚBLICA

1 de 31

BIBLIOTECA DEL CONGRESO

**Digitalización de libros y  
Gestión Documental  
de la Colección patrimonial de la  
Biblioteca del Congreso**

ANEXO TECNICO

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>2 de 31</b>

## PRESENTACIÓN

El presente documento denominado “Digitalización de libros y gestión documental de la colección patrimonial de la Biblioteca del Congreso”, constituye una guía para implantar un proceso que permita digitalizar los libros disponibles en la biblioteca, posteriormente almacenar sus datos de catalogación en una base de datos, indexar a continuación el contenido completo del texto de todos y cada uno de los libros, y por último brindar la disponibilidad de dicha información en una plataforma digital, de modo que se puedan realizar búsquedas profundas en el contenido de los mismos.

El proceso mencionado responde a la necesidad que tiene la Biblioteca del Congreso de organizar, estructurar y facilitar la consulta de los documentos y libros bajo su custodia.

La Biblioteca del Congreso representa un capital histórico para el país, pues allí reposan, entre otros, más de 60 mil volúmenes de la biblioteca de Juan Lozano y Lozano y otros más de la colección que desde la Independencia fue acumulando el Congreso de Colombia.

Obras históricas que datan de los siglos XVII y XVIII enriquecen el ámbito cultural brindado por la Biblioteca, sin desconocer la existencia de obras más recientes en los campos de la Ciencia Política, el Derecho y la Historia, entre otras disciplinas sociales.

En el segundo nivel, de los tres pisos que se levantan sobre el patio central, reposa el único ejemplar que se conserva de la Constitución de 1821, expedida en la Villa del Rosario de Cúcuta, con el gigantesco libro de actas forrado en cuero.

Es este espacio recogido y tranquilo en el que se condensa una mirada a la historia viva del país, y que justifica por su impacto no solo social sino también histórico, su conservación y transformación digital basada en modernas tecnologías tecnológicas, de modo tal que su consulta resulte en un verdadero placer para el espíritu.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>3 de 31</b>

## ANTECEDENTES Y JUSTIFICACIÓN

Enclavada en la zona más histórica y referente mundial en materia turística en América Latina- la localidad de La Candelaria – la Biblioteca del Congreso– se encuentra en el que los Colombianos y extranjeros que la visitan tienen la oportunidad de repasar la historia y fomentar la lectura, haciendo honor a la riqueza que brinda una entidad que ha conservado la historia viva del país, representada en los documentos que han sido puestos bajo su custodia.

El origen de la Biblioteca del Congreso se remonta a la Ley 69 del 31 de diciembre de 1972, que le dio el título de Biblioteca Nacional Especializada. En su interior en la sede en la carrera 6 No 8-94, posee un departamento de archivos y documentos, así como una oficina de referencias legislativas encargada de suministrar a los Senadores y Representantes la información que sustente sus estudios y antecedentes de proyectos de ley o el fortalecimiento de los debates, sobre todo partiendo del hecho que: “La Biblioteca del Congreso suministra a las Comisiones Constitucionales permanentes y especiales, la cooperación y asesoría en el estudio de la agenda legislativa”.

A nivel internacional, este epicentro de la historia de Colombia y del trabajo del Congreso –punto de referencia de la democracia- es comparable a bibliotecas francesas, italianas, alemanas o inglesas.

La Biblioteca está integrada por más de 34 mil volúmenes, donde los visitantes pueden encontrar textos de historia universal, filosofía, ciencias naturales, biografías, derecho, sicología, política, literatura de todo el mundo, arte, libros de viajes, escritos en español, escritos en inglés, francés, italiano, sueco, noruego y portugués, que en su gran mayoría son primeras ediciones, y por tanto poseen un valor histórico innegable.

la Biblioteca del Congreso ofrece servicios adicionales de información propia de una Biblioteca Parlamentaria.

La transformación digital de la Biblioteca es una necesidad, realizar la digitalización de libros de la colección patrimonial y su posterior proceso de gestión documental utilizando modernas técnicas informáticas.

La adopción de un modelo tecnológico y procedimental que preserve los tesoros históricos que guarda en su interior la Biblioteca, se justifica desde múltiples ángulos, pero basta citar entre otros los siguientes: el Acta del Congreso y la Constitución de Cúcuta de 1821; los Censos Generales de Población de 1845 y 1843 y la Constitución Política de Estados Unidos de Colombia, publicada en 1863. Así mismo, destacar las memorias de los Ministerios al Congreso, informes que por sus características y por la importancia que tienen para la historia de Colombia y la labor legislativa, constituyen una colección que requiere tratamientos especiales y que es necesario evaluar, preservar y difundir entre los interesados.

La consulta de estos documentos es restringida por su valor histórico y los riesgos de pérdida, por lo que se requiere un proceso de digitalización que tenga como prioridad la conservación y el uso de nuevas tecnologías para su correcta preservación, consulta y recuperación en caso de ser necesario.

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>4 de 31</b>

## MARCO TÉCNICO

El acceso a la información relevante y pertinente almacenada en la Biblioteca del Congreso, va más allá de una simple búsqueda: es el uso del conocimiento que se deriva de la rica historia del país, sintetizada en los documentos que se guardan en las instalaciones de la Biblioteca.

El proceso, visto de manera global, se descompone en los siguientes pasos:

- Caracterización de los documentos y archivos a ser procesados
- Proceso de digitalización de los documentos, libros y textos, utilizando tecnologías y herramientas que preserven la calidad de los mismos
- Definición de una taxonomía de la documentación procesada, facilitando con ello el acceso y la búsqueda de información por parte de los ciudadanos
- Poblamiento a partir de los documentos originales, de todos los sistemas informáticos de búsqueda y gestión automática de información puestos a disposición
- Utilización de modernas tecnologías que permitan la recuperación de texto de manera ágil y eficiente, con base en herramientas de Inteligencia Artificial
- Generación de funciones avanzadas de búsqueda



## PLATAFORMA COMPUTACIONAL

En el cuerpo del presente documento se define la Plataforma Computacional, como el sistema de software que brinde soporte a todos los procesos indicados, desde la caracterización, pasando por la digitalización documental y subida de información, y la inclusión de un sistema de búsqueda profunda documental. La Plataforma Computacional integra y referencia las mencionadas fases, y posee características integradoras que facilitan la trazabilidad de todos los procesos.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>5 de 31</b>

## CARACTERIZACIÓN DOCUMENTAL

La caracterización documental de la información disponible en la Biblioteca, consiste en la generación previa de datos descriptivos de cada documento, conocidos también como meta datos, los cuales describen cada libro o recurso, facilitando su almacenamiento y posterior consulta.

La caracterización documental se lleva a cabo utilizando bases de datos que almacenan la información descriptiva de los documentos, libros y recursos de la Biblioteca. Esta primera fase permite realizar un inventario físico de los libros y recursos de la Biblioteca, de modo que a través un sistema de búsqueda informático se dispone de información sobre la ubicación, disponibilidad y estado de los recursos que la Biblioteca pone a disposición para su procesamiento según se describe en este documento.

La caracterización debe cumplir con los siguientes parámetros:

- Elaborar la indización para el material digitalizado con cumplimiento de las normas del Archivo General de la Nación y otros que le apliquen
- Catalogar el material digitalizado que incluya alfabetización informacional, ficha técnica del material con título, autor, ubicación, estado, resumen, palabras claves, entre otros. En general que posibilite hacer uso de distintos criterios de búsqueda.

## PROCESO DE DIGITALIZACIÓN

El proceso de Digitalización consiste en transformar los documentos físicos en entidades digitales que serán almacenadas en los repositorios de la institución. Su propósito es permitir su transformación desde lo físico en un recurso de carácter digital, utilizando para ello herramientas basadas en OCR (Optical Character Recognition – Reconocimiento Óptico de Caracteres), como base para su posterior utilización el sistema informática de búsqueda.

Los procesos de reconocimiento de texto basado en tecnología OCR deben estar acompañados por un proceso de conversión de cada documento, en el cual se preserva el estado del libro sin alterar su forma física ni su contenido, a partir del uso de tecnología basada en escáner de tecnología punta.

el cual servirá como base y su posterior búsqueda basada en datos descriptivos (meta datos) y de carácter profundo basado en técnicas derivadas de OCR y algoritmos de organización documental especializados.

El sistema de digitalización contempla las siguientes fases:

- Verificar y catalogar las condiciones del inventario a digitalizar
- Realizar el embalaje, traslado y transporte del material a digitalizar en condiciones que garanticen su conservación fuera de las instalaciones del Senado de la Republica, si es que se va hacer afuera
- Limpieza y desinfección de todo el material entregado
- Reporte por escrito del proceso de limpieza para cada soporte y su estado
- Digitalización del material entregado

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>6 de 31</b>

Estos procesos deben estar debidamente sustentados por software en el ciclo de vida del proceso.

## TAXONOMÍA DOCUMENTAL

La Taxonomía Documental permite establecer una estructura a la cual se podrá recurrir para procesos de búsqueda, y así mismo para la inclusión sistemática de documentos a la plataforma digital.

La definición de la taxonomía documental obedece a dos principios organizativos. El primero de ellos se basa en la estructura lógica de la entidad, y por tanto define las categorías sobre las cuales se organizará de manera estructural el conjunto de documentos digitalizados. Su propósito es el de permitir disponer de un sistema de indexación simple que el buscador avanzado pueda reconocer para realizar búsquedas eficientes y rápidas. La definición de dichas categorías es flexible, y propende por la generación de un entorno en el que se describa el modo en que se encuentra dividida la entidad por secciones y departamentos, y los tipos de documentos que se almacenan en cada una de ellas. Esta primera definición facilita el proceso automático de subida de información a la plataforma, el cual se describa más adelante. Las categorías hacen alusión a los departamentos internos de la entidad o a los procesos que describen el uso y/o almacenamiento de los documentos. La plataforma permite la definición dinámica de estos elementos.

El segundo mecanismo de definición taxonómica se corresponde con la utilización del modelo estándar descriptivo para Bibliotecas, y cuya referencia es el protocolo MARC (Machine Readable Cataloging – Modelo de descripción bibliográfica), elaborado por la Biblioteca del Congreso de los Estados Unidos. Existen varias versiones de MARC, pero en los procesos de implantación de la propuesta se elige MARC 21 por su naturaleza universal. Los registros basados en este estándar permiten la descripción de cada recurso bajo el estándar ISO 2709. Con base en este modelo, la descripción de los campos asociados a cada documento se establece de manera universal, y el tratamiento de software admite aplicativos que cumplan con esta norma.

Los campos de los registros MARC 21 se encuentran descritos en la norma disponible para describir su naturaleza, pero se pueden citar los siguientes: tipo de registro (material textual, material cartográfico, grabación sonora musical, entre otros), nivel bibliográfico (colección, monografía, entre otros), tipo de control archivístico, nivel de codificación, forma de catalogación descriptiva, nivel de un registro multiparte (conjunto, parte con título independiente, entre otros), y todos los demás contemplados en la norma.

Se dispone de software de catalogación para el uso de la norma MARC 21, y así mismo los registros y bases de datos se integran dentro de la plataforma de organización y búsqueda de información, lo cual permite que el sistema se encuentre integrado en todas sus facetas, desde el proceso de caracterización, digitalización, subida de información y procesos de acceso y búsqueda documental avanzada.

## HERRAMIENTA AUTOMÁTICA DE ORGANIZACIÓN

La inclusión en la Herramienta Automática de Organización de la Información está centrada en el uso sistemático de los meta datos (datos descriptivos). Ello hace referencia a los elementos básicos, como título, colección, año, categoría, entre otros, disponibles desde la norma MARC 21, lo que facilita que la subida de información a la plataforma digital responda a las necesidades propias de organización basada en estándares propios del área, como el poder responder al modelo organizativo disponible en el sistema de software de búsqueda avanzado, el cual debe poder permitir búsquedas sobre datos descriptivos y búsquedas sobre el

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>7 de 31</b>

contenido interno de los documentos. En esta segunda alternativa, la plataforma debe poder permitir que los usuarios realicen consultas referidas al contenido específico de los documentos, y la plataforma digital debe poder visualizar los documentos en la página o punto en donde se encuentre dicho contenido, resaltando o haciendo visible la palabra o frase buscada. Este ítem se refiere a lo largo del documento como búsqueda profunda documental.

En esta fase se realiza la subida de documentos a la plataforma, haciendo notar que dentro del sistema se almacenan dos tipos de datos ligados a cada documento: a) Datos descriptivos del documento, b) Contenido interno del documento.

Los datos descriptivos (meta datos) se utilizan para reconocer de manera sistemática los documentos, y desplegarlos según se lo solicite. El contenido interno de los documentos es un valor agregado que se deriva de los procesos avanzados de OCR disponibles en la plataforma de software que da soporte al sistema, a través del cual los usuarios dispondrán de la posibilidad de buscar textos o frases que formen parte del contenido interno y propio de cada documento.

#### SISTEMA AVANZADO DE BÚSQUEDA

La inclusión en el Sistema Avanzado de Búsqueda tiene como propósito permitir la búsqueda sistemática y profunda de información documental, facilitando que los usuarios busquen frases y palabras contenidas en los documentos, desplegando de manera sistemática, rápida y eficiente, dicha información. Esto facilitará los procesos de búsqueda para todos los usuarios.

La plataforma base de software debe permitir que los usuarios realicen solicitudes en lenguaje natural, es decir, en sus propias palabras, o navegando entre las distintas taxonomías disponibles en el sistema, de modo que el motor interno de búsqueda encuentra los documentos que mejor se corresponden con la búsqueda solicitada, despliega la página en la cual se encuentra lo buscado, y lo resalta de modo que resulte visible para los usuarios.

Esta modalidad implica que todos los contenidos internos de los documentos estarán disponibles para búsqueda profunda de los mismos.

El tiempo de respuesta de la Plataforma Computacional ante una consulta de usuario que involucre contenido profundo, debe ser menor a cinco segundos bajo el supuesto de un ancho de banda de Internet considerado normal, tomando como base 10 Megabytes en tiempo actual.

El tiempo indicado para la respuesta no toma en consideración lo que se tarda en desplegar la información propiamente dicha. Esta restricción se refiere a que una vez recibida la consulta en la interfaz del buscador, el sistema de software debe ser capaz de sugerir los documentos más relevantes que contienen en su interior la cadena de texto buscada, y una vez que se elija uno de ellos, la plataforma lo debe obtener desde la ubicación en la web en donde se encuentre almacenado, para desplegarlo a continuación en la interfaz visual del aplicativo.

El tiempo total de búsqueda, en el cual se incluye la obtención de la lista de documentos y el tiempo de obtención del documento seleccionado, así como el texto a resaltar, no debe superar los cinco segundos. El tiempo adicional que tarde la plataforma en desplegar en la interfaz visual la información propiamente dicha,

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>8 de 31</b>

va a depender de la complejidad y el tamaño del documento, y por tanto no se tomará en cuenta dentro del tiempo señalado.

Se valorará positivamente cualquier disminución de tiempo que se obtenga en la plataforma computacional en relación con la obtención de la página y texto asociado en búsqueda profunda.

## HERRAMIENTAS DE DIGITALIZACIÓN

Para el Proceso de Digitalización se utilizan herramientas hardware de tecnología punta, descritos más adelante, a través de las cuales se hace el despliegue digital de los textos almacenados en el repositorio institucional, preservando los elementos físicos de manera eficiente, segura y no invasiva con los documentos.

El proceso de digitalización incorpora fases descritas previamente: caracterización de los documentos a digitalizar, determinación del tipo y herramienta más adecuada para el proceso de digitalización, registro en bases de datos, cronograma de trabajo, aplicación del modelo de digitalización (registrando el proceso realizado), generación del documento digitalizado, inclusión del archivo digital en la plataforma de gestión de software, y gestión de control del proceso realizado, en el cual se determina el cumplimiento de todas las fases descritas (almacenado también en el sistema de información)

Las condiciones de digitalización implican la disponibilidad de las herramientas y plataformas de software requeridas, colocadas a disposición en la Biblioteca para la realización de las tareas indicadas.

Las herramientas y dispositivos físicos de digitalización de documentos, no requieren ningún tipo de manipulación física que altere o estropee en ninguna forma los documentos digitalizados.

Para la Digitalización de los libros no se deberá hacer con descuadernación. Cada libro será preservado en su totalidad, dado que las tecnologías utilizadas de última generación no requieren alterar ni maltratar en lo más mínimo la estructura física del documento. Para ello se emplean sensores electrónicos y dispositivos laser que realizan la labor de digitalización de modo automático y controlado

Resolución óptica (300 dpi), opcional 400 ppp.

Tamaño del libro máximo: 35.5 cm x 34 cm.

Grosor máximo del libro: 15 cm.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>9 de 31</b>

Profundidad de color y Tipos de imágenes

36-bit, color, escala de grises, b & w

En la medida en que se va digitalizando cada página, se debe permitir la visualización en pantalla de la hoja y facilitar un proceso posterior de edición, mediante el cual se asegura una alta calidad en el resultado final.

Se debe poder contar con alternativas para la digitalización de libros y documentos de tipo menos crítico (a evaluar por la Biblioteca), y que permitan una mayor agilidad para el proceso. Las características de resolución deben adaptarse a las ya expuestas previamente.

Se debe disponer de soporte para digitalización de planos y grandes formatos.

Los formatos de salida de los documentos deben obedecer a los formatos más comunes del mercado, PDF, JPG, JPEG, jpeg2000, tiff, tiff G4, png, gif, bmp, pdf (capa incluido OCR), XML, DjVu.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>10 de 31</b>

## ALCANCES

### VOLUMEN DE INFORMACIÓN A PROCESAR

El número de documentos a ser procesados en la plataforma es de hasta 600.000 páginas.

### INTEGRACIÓN DE PROCESOS EN LA PLATAFORMA COMPUTACIONAL

La **PLATAFORMA COMPUTACIONAL**, que representa la base del proceso de desarrollo e implantación, debe incorporar el flujo de trabajo de las siguientes funcionalidades, siendo preferente que todos los procesos y sus resultados se elaboren e integren en una única interfaz de uso. Las funcionalidades incorporadas en el flujo de trabajo de la **PLATAFORMA COMPUTACIONAL** son las siguientes:

**Permitir la generación de taxonomías y categorías de almacenamiento**

**Permitir el almacenamiento de los documentos a partir de los originales digitalizados**

**Realizar la conversión automática de los documentos a una versión que permita las búsquedas de información. En caso de que el documento original no posea dicha estructura, la plataforma computacional debe realizar de manera automática dicha conversión**

**Permitir la visualización del contenido de los documentos a partir de las estructuras taxonómicas definidas**

**Permitir presentar la página que contiene información buscada, resaltando las palabras o frases introducidas por el usuario en la caja de búsqueda. Esta respuesta se corresponde con el concepto de “busca profunda” definida en este documento.**

**Brindar estadísticas apropiadas de la información desplegada en la interfaz visual de la herramienta.**

### DIVULGACIÓN Y PUESTA A PUNTO

En cuanto a la divulgación y puesta a punto, se tiene lo siguiente:

- Entregar e instalar el servidor informando las especificaciones técnicas requeridas que permitan el adecuado almacenamiento, administración, gestión y uso del material digitalizado del Senado de la Republica con las debidas garantías.
- Entregar e implementar el software informando especificaciones y marca para el almacenamiento, la administración y gestión del material digitalizado del Senado de la Republica, debidamente licenciado.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>11 de 31</b>

- Entregar e instalar el Storage y librería automática LTO6, informando especificaciones y marca para el almacenamiento, la administración y gestión del material digitalizado o condiciones del CLOUD.
- Desarrollar y entregar plataforma amigable con el usuario para el acceso del material digitalizado del Senado de la Republica
- Entregar manuales de administración y operación para la plataforma requerida
- Capacitar al personal que requiera el Senado en la administración y operación de la plataforma en un mínimo de 60 horas y entregar un certificado de capacitación por el mismo tiempo.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>12 de 31</b>

## OBJETIVOS

### OBJETIVO GENERAL

Caracterizar la colección de libros y recursos que hacen parte de la Colección Patrimonial del Senado de la República, como base para la digitalización y almacenamiento en el repositorio institucional con base en una descripción taxonómica de los mismos, y la posterior inclusión de los archivos digitales en la Plataforma Computacional, lo que facilita y permite realizar consultas en lenguaje natural o a través de los datos descriptivos de los documentos, generando respuestas sobre los ítems buscados y así mismo la obtención de información referida al contenido profundo de los documentos.

### OBJETIVOS ESPECÍFICOS

Los objetivos específicos del proyecto de digitalización de la Colección Patrimonial del Senado de la República se presentan a continuación:

- Caracterización de los documentos y archivos, tanto físicos como digitales
- Generación de la estructura de datos descriptivos (metadatos) correspondientes a los documentos a ser digitalizados
- Digitalización de los documentos utilizando herramientas y software basado en OCR y técnicas no invasivas documentales
- Realizar el poblamiento de la plataforma computacional con los documentos convertidos en archivos digitales, garantizando que el contenido profundo de los documentos esté disponible para el sistema
- Estructuración de un modelo taxonómico asociado al sistema de la Biblioteca bajo el estándar MARC 21, y un modelo de catalogación estratégica definida en la institución, cuyo propósito es facilitar el acceso eficiente a la información
- Puesta a punto de la herramienta computacional, la cual incorpore los archivos digitales, preparando el entorno visual del sistema para los procesos de búsqueda tanto con base en datos descriptivos como de búsqueda profunda.
- Utilización de una plataforma de gestión documental orientada a bibliotecas, preferiblemente open source, cuya característica es integrar funcionalidades que validen los mecanismos de caracterización, subida automática de información y búsqueda basada en datos descriptivos de los documentos.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>13 de 31</b>

- Los documentos se almacenan siguiendo estándares de conservación y acceso por encriptación segura de su contenido, garantizando su contenido y acceso seguro. Se debe garantizar como mínimo el cifrado AES-256.
- La plataforma será utilizada de acuerdo con los roles y esquemas de acceso definidos por la institución, de modo que los documentos tendrán niveles de visibilidad público y privado, dependiendo del grado de importancia o confidencialidad requerida. Se debe establecer la categoría de visibilidad en la institución, la cual determinará dicho nivel cuando alguno de los documentos tuviera carácter reservado, aspecto que competiere determinar a la institución.
- Documentar los procesos desarrollados
- Realizar la trazabilidad sobre todos los procesos ejecutados

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>14 de 31</b>

## FASES DE IMPLANTACIÓN

Se requiere para la fase de implantación disponer de infraestructura física y de software adecuada. Se procede a describir dichos elementos.

### INFRAESTRUCTURA HARDWARE

El modelo de hardware planteado se orienta a ser implantada sobre los servidores de la institución, acorde con los requerimientos del software y los repositorios de datos.

Una alternativa a ser presentada es una solución CLOUD cuya característica es permitir el trabajo concurrente y seguro de los web services desarrollados en la sección de software. Debe indicarse el alcance, costo y forma de pago.

El costo se asocia al alcance indicado a continuación, y para su valor y forma de pago se hace referencia a una de las plataformas existentes, entre las cuales destaca Google Clouds y Amazon.

El alcance técnico se describe a continuación:

La interacción entre ambos sistemas se realizará a través de Web Services.

El modelo hardware propuesto se compone de los siguientes elementos:

- Cloud DNS. Para reconocimiento de IPs asociadas a los procesos
- VPC. Virtualizador del servicio de red
- LB. Balanceador de carga. A través de él se logra establecer equilibrio en los proceso de conversión de documentos
- Vmi. Máquinas virtuales asociadas a los procesos de conversión.
- Persistent Storage, Crypted. Almacenamiento persistente de los datos convertidos, con protocolos de encriptación que garanticen la seguridad global de la plataforma

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>15 de 31</b>

- Database. Módulo de almacenamiento persistente general, para el mantenimiento histórico y de acceso instantáneo a los datos, con base en el uso inteligente de los patrones de búsqueda
- Replica. Espejo de los datos almacenados, para generar un nivel de seguridad adicional

Para mejorar aún más la concurrencia se utilizan “zonas de disponibilidad”, lo cual brinda un mejor desempeño de la plataforma.

#### INFRAESTRUCTURA SOFTWARE

Las herramientas tecnológicas deben ser operadas bajo las siguientes consideraciones técnicas:

- Sistema Operativo: Linux
- Servidor: Apache
- Motor de base de datos: Oracle
- Lenguajes de desarrollo: PHP, JavaScript, JQuery, entre otros.

Los productos desarrollados deben ser servicios web cuya característica es la de permitir que se puedan integrar de manera transparente a los sistemas de información del Senado de la República.

Los servicios web se basan, inicialmente, en componentes RESTfull. El objetivo del desarrollo es permitir que se genere interacción transparente con los aplicativos de la institución, lo cual permite que el uso de las respuestas obtenidas (usualmente JSON y XML) puedan ser procesadas y utilizadas en el modelo XHTML, garantizando con ello la transparencia de los procesos, y así mismo permitiendo que la presentación de los resultados sea independiente de los desarrollos de las herramientas tecnológicas.

Los procesos de inteligencia artificial utilizan TensorFlow como base para el desarrollo de redes neuronales convolucionales, y para la gestión de los sistemas expertos se recurre a CLIPS, el cual se formaliza a través de commonKads como metodología de desarrollo.

Todos los elementos indicados operan bajo una plataforma que en su origen se compone de herramientas open source: JQuery, node.js, JavaScript, HTML5, CSS3.

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>16 de 31</b>

## ENTREGABLES

Nro.	Entregable
1	Plataforma Computacional de despliegue integrado
2	Taxonomía para el poblamiento, basada en MARC 21
3	Página web del producto
4	Herramientas de digitalización disponibles
5	Software de gestión de caracterización y seguimiento
6	Manual de usuario
7	Manual técnico
8	Informe integral del proceso desarrollado

## REPOSITORIOS INSTITUCIONALES

Los repositorios de datos utilizados por la plataforma se almacenan en servidores que cumplen de manera estricta con los requisitos formales de seguridad, duplicación de la información, y acceso permanente. Se debe ofrecer almacenamiento externo basado en CLOUD, como alternativa a los servidores propios de la institución.

## ANALÍTICA DE DATOS

Se debe realizar la incorporación adicional de un modelo de dashboards para el despliegue de la gestión documental, acceso a la plataforma, niveles de consulta, tendencias de uso de la información, tópicos más buscados, relación de documentos que forman parte de cadenas de consulta, entre muchas otras.

El dashboard de analítica de datos es una opción que utiliza cubos de decisión y técnicas de computación blanda que brindan análisis de tendencias basadas en redes neuronales, lógica difusa y sistemas expertos, con base en TensorFlow y CLIPS, entre otros.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>17 de 31</b>

## ESTÁNDARES DE GESTIÓN DOCUMENTAL

El contratista deberá cumplir con los estándares de la gestión documental para bibliotecas.

Se requiere adhesión al protocolo Dublin Core. El modelo de metadatos Dublin Core ha sido elaborado y auspiciado por la DCMI (Dublin Core Metadata Initiative), la cual es una organización dedicada a fomentar de manera amplia y extensa la adopción de los estándares interoperables de los metadatos, y así mismo a promover el desarrollo de los vocabularios especializados de metadatos con el objetivo de describir recursos que permitan la aparición de sistemas más inteligentes relativos al descubrimiento del recurso.<sup>1</sup>

Las implementaciones usuales de Dublin Core utilizan generalmente XML, aunque en algunos casos pueden orientarse a JSON, y se basan en el Resource Description Framework. Dublin Core definido por ISO en su norma ISO 15836 del año 2009, y la norma NISO Z39.85-2012

Adicionalmente, es necesario que se acoja la norma MARC 21, como base para la caracterización del sistema documental.

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>18 de 31</b>

### VALOR DE LA PROPUESTA

Se debe incluir el valor de la propuesta, acorde con el siguiente modelo:

	<b>Cantidad</b>	<b>Unidad de medida</b>	<b>Valor unitario sin IVA</b>	<b>IVA</b>	<b>Valor unitario con IVA</b>	<b>SUBTOTAL VALOR CON IVA</b>
<b>DIGITALIZACION</b>	<b>600.000</b>	<b>Páginas</b>				
<b>IMPLEMENTACION INFRAESTRUCTURA TECNOLOGICA</b>	<b>1</b>	<b>Unidad</b>				
<b>SOFTWARE DE CATALOGACIÓN E INDIZACION</b>	<b>1</b>	<b>Unidad</b>				
<b>SERVICIOS, PUESTA EN MARCHA (Incluye soporte y mantenimiento)</b>	<b>1</b>	<b>Año</b>				
<b>VALOR DE LA PROPUESTA</b>						

	ANEXO TÉCNICO	Página
	DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO	19 de 31
SENADO DE LA REPÚBLICA		

## EQUIPO HUMANO

Para una adecuada ejecución del contrato se requiere un equipo de trabajo mínimo destinado que cumpla como mínimo con los siguientes perfiles:

Cargo	Nro.	Responsabilidad	Dedicación	Formación Profesional	Experiencia profesional
Administrador del proyecto	1	Liderar el proceso de implantación completa del proyecto	Tiempo completo	Ingeniero afín con el área de los sistemas y la computación, o con la Ingeniería Eléctrica o Electrónica, o afín con Ingeniería Industrial o Administración, preferiblemente con formación certificada en PMP y/o especialización o maestría en arquitectura o desarrollo de software	Experiencia certificada de al menos 3 años durante los cuales participó en dirección y/o implantación, de proyectos de sistemas de software
Ingeniero Líder de implantación del Proyecto	1	Dirección del proceso de implantación de la Plataforma Computacional	Tiempo completo	Ingeniero afín con el área de los sistemas y la computación, y preferiblemente con especialización o maestría en arquitectura de software, o desarrollo de software	Experiencia certificada de al menos 2 años durante los cuales participo en la implantación de productos de software
Ingeniero de soporte	1	Soporte de la plataforma integrada	Tiempo completo	Ingeniero afín con las áreas de sistemas y computación	Experiencia certificada de al menos 2 años durante los cuales participó en

	ANEXO TÉCNICO		Página
	DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO		20 de 31
SENADO DE LA REPÚBLICA			

					soporte de software
Programador	2	Desarrollo de módulos requeridos por los usuarios y la plataforma	Tiempo completo	Ingeniero afín con las áreas de sistemas y computación, preferiblemente con formación certificada en desarrollo web	Experiencia certificada de 2 años durante los cuales participó en desarrollo web, bases de datos y manejo de frameworks
Ingeniero Web Master	1	Desarrollo sitio web y control de servidores	Tiempo completo	Ingeniero afín a las áreas de sistemas y computación, preferiblemente con formación en redes y formación en desarrollo web	Experiencia certificada de 2 años durante los cuales participó en administración de redes y construcción de portales web
Experto en sistemas de digitalización	1	Programación Proceso de digitalización	Tiempo completo	Ingeniero o Tecnólogo afín con el área de la electrónica o los sistemas de cómputo, preferiblemente con especialización y/o formación en sistemas digitales	Experiencia certificada de 2 años durante los cuales participó en actividades en el campo de los sistemas digitales, la electrónica o los sistemas aplicados al área del hardware,
Operario del sistema software	2	Tareas manuales o semiautomáticas de subida de información	Tiempo completo	Tecnólogo o Técnico afín al área de los sistemas y la computación	Experiencia certificada de 2 años durante los cuales participó en actividades en el uso técnico de sistemas de software
Operarios del sistema de digitalización	2	Proceso manual de digitalización	Tiempo completo	Técnico afín al área de sistemas y/o	Experiencia certificada de 2 años durante los cuales

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>21 de 31</b>

				<b>instrumentación y/o electrónica</b>	<b>participó en las áreas de sistemas digitales y/o electrónica y/o, sistemas y computación</b>
--	--	--	--	--	---

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>22 de 31</b>

## CONSIDERACIONES TÉCNICAS

En términos específicos, las consideraciones técnicas a cumplir son las siguientes:

<b>SOFTWARE PARA LA GESTIÓN INTELIGENTE DE INFORMACIÓN DISTRIBUÍDA</b>
Requerimientos mínimos y especificaciones:
<b>RESTRICCIONES ESPECÍFICAS</b>
Como mínimo el sistema de información debe poder indexar los siguientes tipos de repositorios: <ul style="list-style-type: none"> <li>• Intranet</li> <li>• Servidores web que manejen al menos los protocolos: HTTP 1.0 y HTTP 1.1</li> <li>• Portales administrados por la Institución</li> <li>• Sistemas de archivos compartidos, mínimo los siguientes: (Microsoft, Linux, Unix, Mac)</li> <li>• Bases de datos (Oracle, SQL Server, MYSQL)</li> </ul>
Las taxonomías permiten establecer las rutas específicas de búsqueda, cada una de ellas definida a partir de una URL en concreto.
El servicio debe brindar la conexión con fuentes de información en tiempo real para incluirla como parte de los resultados.
El servicio debe organizar los resultados derivados de la búsqueda, utilizando diferentes criterios: taxonomía, fecha, relevancia, entre otros.
El sistema de software debe brindar ayudas textuales al momento de digitar la búsqueda.
El sistema de software debe permitir la generación automática de búsquedas relacionadas con base en la taxonomía definida.
El sistema de software debe realizar sugerencias en el campo de búsqueda con base en las actividades realizadas previamente y como consecuencia de la estructura interna del sistema, como diccionarios de datos y componentes taxonómicos de la plataforma.
El sistema de software debe permitir agrupar resultados a partir de filtros escritos en lenguaje natural.
El sistema de software debe permitir la búsqueda en documentos, ubicando para ello los documentos que cumplen con la cadena de búsqueda, o las porciones de texto (dentro del documento) que se caracterizan por estar en consonancia con la cadena buscada.

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>23 de 31</b>

<p>El sistema de información debe poder indexar contenido en archivos de aplicaciones comunes. Por ejemplo:</p> <ul style="list-style-type: none"> <li>• Acrobat (PDF)</li> <li>• Excel (En sus distintas versiones)</li> <li>• Word (En sus distintas versiones)</li> <li>• PowerPoint (En sus distintas versiones)</li> <li>• WordPerfect</li> <li>• RTF</li> <li>• HTML</li> <li>• XML</li> <li>• Dbf, dbt</li> <li>• msg</li> <li>• Entre otros.</li> </ul>
<p>El servicio debe soportar la búsqueda de documentos creados con cualquier versión de Microsoft Office, a la fecha de la instalación de la plataforma.</p>
<p>El sistema de software debe permitir indexar diversos sitios web en el propio dominio de la organización u otros dominios y subdominios administrados por la Institución, sin importar el tipo de servidor web utilizado o aplicación usada para la creación del contenido.</p>
<p>La plataforma debe poder realizar búsquedas en sitios web externos a la entidad.</p>
<p>La plataforma debe poder indexar información en formato XML y JSON.</p>
<p>El sistema de información debe poder escalar en cuanto al número de documentos indexados sin ningún cambio del hardware provisto inicialmente.</p>
<p>El sistema de software no debe indexar contenido duplicado o con vínculos múltiples.</p>
<p>El sistema de software debe indexar el texto al interior del documento, generando objetos HTML relacionados.</p>
<p>El sistema de software debe poder realizar búsquedas al interior de las bases de datos, en todos los campos relacionados con la taxonomía definida.</p>

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>24 de 31</b>

<p>El sistema de software deberá poder soportar actualizaciones y crecimiento en el número de documentos a indexar (sin la necesidad de cambios de hardware) hasta para al menos 20.000.000 de documentos.</p>
<p>El sistema de software deberá proveer mecanismos de autenticación única.</p>
<p>El sistema de software debe proveer un mecanismo de administración (back-end) de usuarios, artículos y taxonomías, sin costo adicional.</p>
<p>El sistema de software debe permitir la generación de reportes ligados a los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>• Consultas por hora</li> <li>• Consultas por día</li> <li>• Ranking de las palabras y/o artículos más buscados (encontrados y no encontrados)</li> <li>• Frases más buscadas (encontradas y no encontradas)</li> </ul>
<p>El sistema de software presentará el documento simplificado como objeto de selección, de modo que el usuario pueda decidir si continúa o no con el proceso de despliegue de la información.</p>
<p>El sistema de software debe poder indexar bases de datos relacionales, en campos de texto y otros tipos de datos, como por ejemplo: Oracle, SQL Server y MySQL, entre otros. El número total de bases de datos a indexar no estará limitado más que por el número de licencias disponibles.</p>
<p>Debe permitir la búsqueda de contactos o personas dentro de la organización, utilizando servidores LDAP o Active Directory.</p>
<p>La plataforma deberá soportar mecanismos de autenticación tales como LDAP y formularios en HTTP.</p>
<p>El sistema de software debe estar en capacidad de realizar búsquedas en al menos dos idiomas: inglés y español</p>
<p><b>RESULTADOS GENERADOS COMO RESPUESTA A CONSULTAS DE USUARIO</b></p>
<p>Una vez definida la estructura taxonómica del sistema de información asociada a cada dependencia, la plataforma de búsqueda elegirá las mejores respuestas ante las solicitudes de los usuarios, navegando por el árbol de conocimiento generado a través de la plataforma.</p>
<p>El sistema de software modifica la importancia de los documentos de acuerdo con las selecciones efectivas que el usuario realice sobre los resultados alternativos generados por las búsquedas.</p>

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>25 de 31</b>

<p>El software clasificará los documentos y recursos utilizando criterios como: fecha de creación de los artículos, repositorio específico de almacenamiento y ubicación en el árbol taxonómico en la plataforma.</p>
<p>El software brindará resultados alternativos con base en el árbol taxonómico de la plataforma, o contenidos relacionados que coincidan con palabras similares, sinónimos y términos contenidos en el diccionario base de la plataforma.</p>
<p>El sistema de software posicionará en los primeros lugares de despliegue de información, los contenidos previamente definidos en la taxonomía, o los que tengan un mayor número de selecciones cuando la búsqueda no sea exacta.</p>
<p>El sistema ajustará la búsqueda al resultado más cercano, a pesar de que el usuario cometa errores ortográficos, con base en las taxonomías definidas y los target documentales.</p>
<p>El buscador permitirá la utilización de filtros utilizando lenguaje natural, de modo que se restrinja el espacio de búsqueda de información</p>
<p>El servicio de software permitirá la creación de categorías de búsqueda (taxonomías) a partir de una estructura relevante para el usuario del sistema, según sus necesidades y a través de herramientas debidamente elaboradas con este propósito.</p>
<p>El sistema debe incluir la generación automática de diccionarios de datos a través de los cuales se establezcan patrones sintácticos que permitan las búsquedas dinámicas relacionadas con las cadenas de búsqueda de los usuarios.</p>
<p>El software debe brindar sugerencias de manera dinámica en el proceso de búsqueda, permitiendo autocompletar el texto a medida que introduce la consulta.</p>
<p>Permitir filtrar los resultados generados acorde con las taxonomías establecidas y los contenidos de los documentos almacenados en los repositorios de datos.</p>
<p>El sistema de software debe permitir la integración con sistemas de RSS de WordPress, contenido de Twitter y Flickr.</p>
<p><b>DESPLIEGUE DE INFORMACIÓN ESTRATÉGICA</b></p>
<p>La información estratégica y relevante que los usuarios determinen como clave para sus procesos misionales, estará disponible en una interfaz separada en el medio de despliegue, de modo que resulte visible y accesible por parte del usuario</p>
<p><b>VISUALIZACIÓN DE CONTENIDOS OBTENIDOS A PARTIR DE LAS BÚSQUEDAS</b></p>

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>26 de 31</b>

El sistema de software desplegará como resultado de la búsqueda, un objeto HTML como hiper enlace al recurso solicitado. Al seleccionar el objeto HTML, se visualizará un documento estándar HTML sobre la interfaz del navegador. El documento tendrá asociado un enlace a la categoría taxonómica correspondiente, de modo que se pueda navegar a través del árbol de búsqueda asociado. En el caso de documentos estándar (PDF, Excel, TXT, Word, entre otros), se desplegará el correspondiente objeto HTML, resaltando las palabras buscadas, y facilitando la navegación entre los diferentes contenidos. Para documentos PDF, y de manera opcional, se desplegará dicho documento utilizando los añadidos de software que traen los navegadores, en caso de que estos se encuentren presentes. Los vídeos serán abiertos como resultado de búsqueda, a partir de la disponibilidad de añadidos de software para tal fin en el navegador del usuario.

#### **CARACTERÍSTICAS TÉCNICAS DE LA PLATAFORMA**

El sistema debe poseer un mecanismo que permita convertir documentos PDF o imágenes a objetos HTML, de manera directa y automática, sin intervención manual. Bastará con introducir la dirección URL del documento, para que se obtenga de manera automática el objeto HTML a ser desplegado en la pantalla de resultados.

El sistema de software debe proveer un mecanismo de llenado de información, de modo que tales herramientas permitan insertar contenido y artículos relativos a la taxonomía de la dependencia, preferencialmente de manera automática o semiautomática.

El sistema de software se podrá instalar en dos modalidades: 1) Sobre el servidor de la Institución, utilizando para ello un sistema de instalación y configuración. 2) Directamente sobre el servidor del Proveedor, el cual entregará las credenciales adecuadas y habilitará el software para ello. La Institución decidirá el modo más adecuado según sus necesidades.

Se debe incluir para el servidor de aplicaciones los siguientes elementos de tolerancia a fallos, como mínimo:

- Discos en RAID
- Segunda fuente (redundancia)

El sistema debe ser escalable en cuanto al modelo de búsquedas a partir de un esquema basado en configuración por software, bajo el supuesto de que la plataforma hardware esté configurada en cuanto al gestor de base de datos y capacidad de los repositorios de datos.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>27 de 31</b>

<p>El sistema debe permitir la administración de usuarios, estableciendo sus perfiles y niveles de seguridad, en una base de datos corporativa o bajo el protocolo que la Institución defina para tal propósito.</p>
<p>El sistema de software debe manejar información en tres niveles: 1) Información global, 2) Información privada al usuario, 3) Información disponible únicamente para la dependencia. Para los tres casos, el sistema de búsqueda desplegará únicamente la información que se corresponda con estos criterios. En ningún caso se desplegará información a la que no se tenga derecho. La plataforma contará con un sistema de estadísticas de búsquedas, tanto de manera grupal como individual.</p>
<p>El administrador del sistema podrá monitorear la plataforma a través de la interfaz web.</p>
<p>El sistema de software debe permitir acceso remoto mediante el protocolo SSH, de modo que desde consola (o herramientas asociadas) se puede brindar soporte técnico por parte del fabricante del sistema.</p>
<p>El sistema de software debe ser compatible con las plataformas más utilizadas, entre las que se incluye como mínimo: sistemas operativos Windows y Linux, AIX, Solaris, MAC, entre otros.</p>
<p><b>SOLUCIÓN REQUERIDA EN LA INSTITUCIÓN</b></p>
<p>El software debe implantarse sobre un sistema que brinde y soporte el hardware, software y procesos requeridos para su correcto funcionamiento y escalabilidad, lo cual incluye (pero no se limita) a:</p> <ul style="list-style-type: none"> <li>• Servidor (es)</li> <li>• Sistemas operativos</li> <li>• Bases de datos</li> <li>• Portales de administración</li> <li>• Actualizaciones liberadas por el fabricante</li> <li>• Instalación</li> <li>• Configuración</li> <li>• Operación</li> </ul>
<p>Todos los elementos mencionados deben disponer de las licencias y el soporte necesario durante un año (12 meses), para garantizar la continuidad del servicio.</p>
<p>La garantía de funcionamiento de las herramientas tecnológicas involucradas en la prestación del servicio debe incluir cambios en el hardware en caso de daños o actualización de versiones liberadas por el fabricante sin costo adicional para la entidad.</p>

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>28 de 31</b>

<p>La vigencia de la garantía de los productos que lo requieran, debe ser de al menos un año (12 meses).</p>
<p>El servicio incluye la entrega en el sitio designado por la Institución, y así mismo la provisión de los equipos necesarios para la prestación del servicio.</p>
<p>Las actualizaciones de la plataforma, cuando estas se presenten, no tendrán ningún costo extra para la entidad en el marco del tiempo de vigencia del contrato</p>
<p>La única conexión a los repositorios de información será a través de puertos de red Ethernet, facilitando con ello la integración con la infraestructura de la entidad</p>
<p>La solución informática debe poder escalar a la capacidad para tolerancia a fallos y la continuidad de la operación en caso de requerirse.</p>
<p>Las labores de mantenimiento durante el periodo de licenciamiento son las siguientes:</p> <ul style="list-style-type: none"> <li>• Mantenimiento correctivo</li> <li>• Mantenimiento preventivo (por lo menos una vez al año).</li> <li>• Actualizaciones de firmware.</li> <li>• Actualizaciones de software.</li> <li>• Soporte técnico sobre la solución informática por un año en sitio (cuando sea requerido) y telefónico, con una disponibilidad 5x8 en el mismo día, o como máximo al día hábil siguiente</li> </ul>
<p><b>SERVICIOS PROFESIONALES BRINDADOS POR EL PROVEEDOR</b></p>
<p>El proveedor del sistema de software deberá brindar apoyo a la Institución para que la solución informática sea compatible con la plataforma de base tecnológica de la entidad</p>
<p>El proveedor brindará asesoría en cuanto a las dudas de carácter técnico que se presenten durante el proceso de ejecución e implantación del sistema.</p>
<p>El proveedor construirá una interfaz por dependencia, y hasta máximo cinco (5) dependencia en el marco del presente contrato. Cada interfaz dispondrá de todos los elementos descritos en este documento de referencia.</p>
<p>Durante la ejecución del presente contrato, el proveedor administrará de manera remota o presencial, de ser necesario, la plataforma de búsqueda realizando cambios en la configuración, reportes del sistema, mantenimiento de software, y tareas correlacionadas con el normal funcionamiento y operatividad de la plataforma.</p>
<p><b>ENTREGABLES (Adicionales a los establecidos en los alcances de presente documento)</b></p>

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>29 de 31</b>

Plan detallado de implantación y cronograma de actividades.
Evidencia de la incorporación de las fuentes de información disponibles en la Institución a los repositorios institucionales y su inclusión en la taxonomía de la institución.
Reportes de resultados, al finalizar cada mes de actividades.
Entregar el documento en el que se indique la suscripción de las licencias y los derechos del servicio de soporte por un (1) año a nombre de la Institución, para todos los componentes de hardware y software.
<b>TIEMPO DE ENTREGA</b>
La Plataforma Computacional integradora de todos los procesos debe ser instalada y configurada en las instalaciones que la Institución determine, en un plazo no superior a 15 días calendario, contados a partir del perfeccionamiento del Contrato en todas sus partes, previo registro presupuestal y así mismo de la aprobación de pólizas correspondientes. Este ítem hace referencia al sistema de cómputo de integración descrito previamente.
Los reportes mensuales correspondientes a la operación de la plataforma se enviarán a la Institución durante los primeros cinco (5) días hábiles del siguiente mes que es objeto del reporte.
El plan detallado de trabajo debe entregarse en un plazo no mayor a cinco (5) días calendario a partir del perfeccionamiento del contrato.

#### LICENCIAMIENTO Y SOPORTE FORMAL

**Se brinda licencia de uso del producto sin restricción de tiempo, lo cual implica que la licencia no caducará en ningún sentido en el marco especificado en este documento. Por tanto, la Biblioteca dispondrá de todas las funcionalidades del producto debidamente licenciadas por el proveedor, sin que caduque la licencia en ningún momento del tiempo. Las mejoras que cambien el funcionamiento de la Plataforma Computacional implantada y aprobada entre las partes, requerirá de un nuevo acuerdo y de un nuevo licenciamiento.**

Los datos utilizados por el Contratista serán únicamente aquellos que estén disponibles en el repositorio de datos, o los que formen parte de los portales institucionales o de acceso público. Por lo ya expresado, es responsabilidad exclusiva de la institución el garantizar la calidad, pertinencia y validez de la información desplegada.

	<b>ANEXO TÉCNICO</b>  <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>30 de 31</b>

Para validar que la información almacenada en el sistema sea fidedigna, se dispone de un sistema de usuarios y roles que garantizan que el acceso se realice a través del modelo de seguridad que el Senado de la República determine con antelación de manera concertada con el contratista.

Por último, se debe indicar que la confidencialidad de los datos se debe manejar directamente desde la Institución, pues el contratista no realiza ningún tipo de manipulación sobre ellos.

#### PROCESO DE CAPACITACIÓN

Se brindará capacitación a todos los funcionarios de la institución en el manejo de la Plataforma Computacional de software, hasta un total de tres cubriendo todas las dependencias internas de la Biblioteca. Se complementará la capacitación con un manual de usuario y un vídeo que ilustre el manejo del sistema.

La capacitación toma en cuenta el perfil de los usuarios, adecuándolos al rol a ejecutar: administrador, usuario final.

#### MANTENIMIENTO Y SOPORTE

El soporte y mantenimiento de la aplicación, una vez realizada la instalación, configuración y poblamiento del sistema sobre el cual opera la plataforma será realizada por el Contratista. El soporte se brinda por un período durante todo el período de contratación, y su periodicidad es permanente, para lo cual la empresa dispondrá de un funcionario con la disponibilidad de tiempo pertinente para dichas actividades.

#### CARACTERÍSTICAS TÉCNICAS

En atención al proceso de implantación y seguimiento, el proveedor del producto se compromete a:

- Disponer de las licencias y soporte que garantice el correcto funcionamiento de la plataforma durante el primer año, sin costo adicional
- La continuidad del servicio incluye actualización de la plataforma en caso de cambios tecnológicos que así lo requieran, en el caso de ajuste de las herramientas que constituyen el sistema
- El proceso de implantación se hará directamente sobre las plataformas de la institución

	<b>ANEXO TÉCNICO</b> <b>DIGITALIZACIÓN DE LIBROS Y GESTIÓN DOCUMENTAL DE LA COLECCIÓN PATRIMONIAL DE LA BIBLIOTECA DEL CONGRESO</b>	<b>Página</b>
	<b>SENADO DE LA REPÚBLICA</b>	<b>31 de 31</b>

- El proveedor administrará de manera remota o presencial la plataforma durante el período de garantía de la misma, la cual es de un año contado a partir de la instalación del aplicativo
- Se brindarán reportes mensuales del proceso de implantación
- Se brindará un reporte final del proceso de implantación

**KEYLA MENESES.**

**Biblioteca del Congreso.**