

**Muyal Alwa: Repositorio de servicios,  
datos, contenidos y resultados producidos  
por sistemas de e-Salud.**



**Proyecto: 41756**

**Responsable técnico:** Dr. José Luis González  
Compeán, Profesor investigador, Cinvestav  
Tamaulipas

## Autorizaciones

Fecha: 20/12/2022	Publicación	Versión 1.0
Elaboró	Alfredo Barrón Rodríguez	20/12/2022
Revisó	Dr. Hugo German Reyes Anastacio	20/01/2023
Autorizó	Dr. José Luis González Compeán	

## Resumen

En este documento se describe **Alwa**, un repositorio de servicios de intercambio, preservación, aseguramiento y análisis de datos clínicos de la plataforma, así como sistemas de e-salud publicadas por la misma. Este servicio se compone de repositorios localizables, accesibles, interoperables y reutilizables (FAIR, por sus siglas en inglés), así como base de datos producidas por los sistemas/servicios de e-salud para el intercambio de contenidos. Este servicio permite que las instituciones publiquen datos clínicos para su consumo interno o por parte de la comunidad científica y/o profesionales de salud.

## Índice

Resumen.....	2
Índice de figuras .....	3
Introducción .....	4
Principios de diseño del modelo de construcción de Alwa para repositorios FAIR y base de datos ..	4
Principios FAIR.....	4
DSpace.....	5
Comunidades.....	6
Colecciones.....	6
Ítems.....	6
Painal .....	6
Organizaciones .....	7
Catálogos .....	7
Archivos.....	8
Alwa.....	8
Comunidades y organizaciones.....	8
Colecciones y catálogos.....	9
Ítems y archivos.....	9
Implementación del prototipo Alwa .....	9
Bibliografía .....	11

## Índice de figuras

Figura 1 Organización de metadatos de DSpace.....	6
Figura 2 Organización de datos en Painal .....	7
Figura 3 Organización de Alwa con la integración de DSpace y Painal.....	8
Figura 4 Interfaz gráfica de Alwa con sus respectivas comunidades y organizaciones .....	9
Figura 5 Acceso a los catálogos y colecciones de los servicios disponibles .....	10

## Introducción

La gestión de datos de investigación es el proceso activo de manejo de datos generados en una investigación. Esta gestión se realiza en forma continua y cubre todas las decisiones relacionadas con la gestión de datos a largo de su ciclo de vida, comenzando con la etapa de planificación de la investigación y abarcando su ejecución, la diseminación de los resultados y la preservación de los conjuntos de datos de forma que estos datos sea auténticos, precisos, completos y fiables, y se mantengan accesibles y reutilizables a lo largo del tiempo. Algunos aspectos que trata la gestión de datos de investigación son i) el aseguramiento de la calidad de datos, ii) la creación o captura de metadatos, iii) el almacenamiento y respaldo, iv) la integridad, seguridad y privacidad de los datos, v) el depósito de los datos y su perseveración a largo plazo y vi) el manejo de derechos asociados. Los datos han sido desde siempre la materia prima fundamental para la investigación y el avance del conocimiento. El avance tecnológico de las últimas décadas ha tenido impacto notorio en el volumen de información existente, fenómeno que ha sido caracterizado por el diluvio de datos, lo cual han surgido los siguientes términos e-ciencia, e-investigación y e-salud. Donde este proyecto se enfoca en los datos generados por sistemas de e-salud, que es la salud digital que se refiere al uso de las tecnologías de la información y comunicaciones a modo de herramientas, que se emplean en el sector sanitario en materia de prevención, diagnóstico, tratamiento, seguimiento, así como la gestión de salud mejorando la eficacia de este entorno mencionado.

## Principios de diseño del modelo de construcción de Alwa para repositorios FAIR y base de datos

El repositorio de servicios Alwa se define y está basado en los principios FAIR que permite la compartición de datos abiertos, donde estos principios no solo se centran en los datos, sino que también se esperan ser aplicados en los metadatos, identificadores, software y planes de gestión de datos haciendo que todos estos elementos permitan que los datos sean FAIR.

### *Principios FAIR*

Los principios FAIR para la gestión de datos de investigación son *Findability* (localizable), *Accessibility* (accesible), *Interoperability* (interoperable) y *Reusability* (reutilizable) (Boeckhout, 2018). Estos principios permiten descubrimientos más amplios e innovación de conocimiento basada en la integración y reutilización de datos publicados por la comunidad científica. La ciencia ha comenzado a requerir planes de gestión de datos generados con fondos públicos, permitiendo incluir la noción de “cuidado a largo plazo” de activos digitales con el propósito de ser descubiertos y utilizados para investigaciones posteriores. Los principios utilizados son los siguientes (GO FAIR, s.f.):

- **Localizable** (*Findability*): Los conjuntos de datos deben describirse, identificarse, y registrarse o indexarse de forma clara e inequívoca.

- **Accesible** (*Accessibility*): Los conjuntos de datos deben ser accesibles mediante un procedimiento de acceso claramente definido, idealmente utilizando medios automatizados. Los metadatos deben permanecer siempre accesibles.
- **Interoperable** (*Interoperability*): Los datos y metadatos se conceptualizan, expresan y estructuran utilizando normas comunes publicadas.
- **Reutilizable** (*Reusability*): Las características de los datos y su procedencia describen detalladamente de acuerdo con las normas comunitarias pertinentes, con condiciones de uso claras y accesible.

Las condiciones para aplicar responsablemente los principios FAIR de gestión de datos son:

- **Facilitar el intercambio y la reutilización de datos:** Organización y gobierno de iniciativas puesta en común de datos en comunidades de prácticas FAIR.
- **Tener en cuenta las opciones de diseño:** Explicación, formalización y actualización continua de las normas sobre los datos y metadatos.
- **Respetar el derecho de los participantes:** Desarrollo de marcos y métodos para la privacidad y la protección de datos.

## ***DSpace***

DSpace es uno de los programas de código abierto preferido por las instituciones académicas para gestionar repositorios de archivos (textuales, audio, video, etc.), facilitando su depósito organizándolo en comunidades, asignándoles metadatos y permitiendo su difusión en recolectores y agregadores (Rodríguez-Gairín, 2008). El objetivo de DSpace es coleccionar y organizar la producción intelectual de una institución, describirla y preservar sus objetos digitales en el tiempo. DSpace permite i) facilitar la captura y la inclusión de materiales, incluyendo sus metadatos, ii) facilitar el acceso los mismos metadatos de manera sencilla y iii) facilitar la preservación a un largo plazo de los materiales creados.

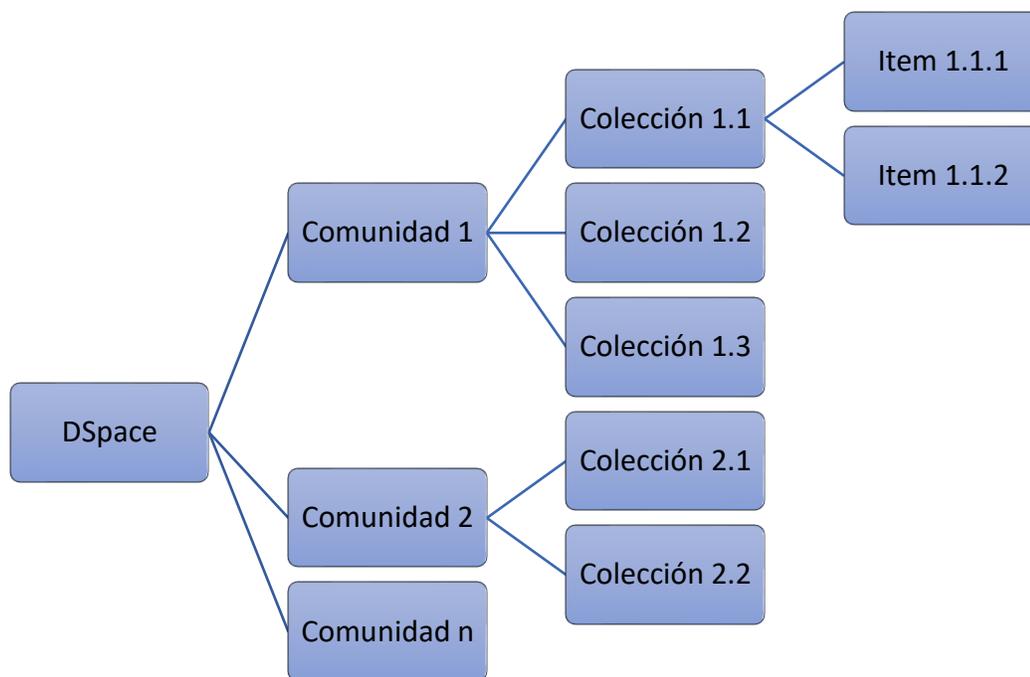


Figura 1 Organización de metadatos de DSpace

La estructura del programa DSpace está compuesta de comunidades, subcomunidades, colecciones e ítems como se presenta en la Figura 1.

### Comunidades

Las comunidades son grupos en las cuales se organiza la información que se agrega a DSpace. Estas comunidades a su vez contienen colecciones o subcomunidades (opcionalmente).

### Colecciones

Las colecciones agrupan ítems que los diferentes miembros agregan dependiendo de sus permisos.

### Ítems

Los ítems se componen de una serie de metadatos y un conjunto de archivos.

### Painal

Painal es un conjunto de servicios y sistemas desarrollados para que las organizaciones de salud y la comunidad científica puedan: i) almacenar, distribuir y localizar sistemas o servicios de procesamiento a través de catálogos de servicios; ii) generar soluciones que permitan brindar rentabilidad costo-beneficio del almacenamiento y transporte de datos; iii) almacenar, publicar y transmitir repositorios de datos de manera local (intra-institucional) y federada (inter-institucional) utilizando un modelo de publicación/suscripción interconectando la infraestructura de TI privada con servicios de nube (pública o híbrida)

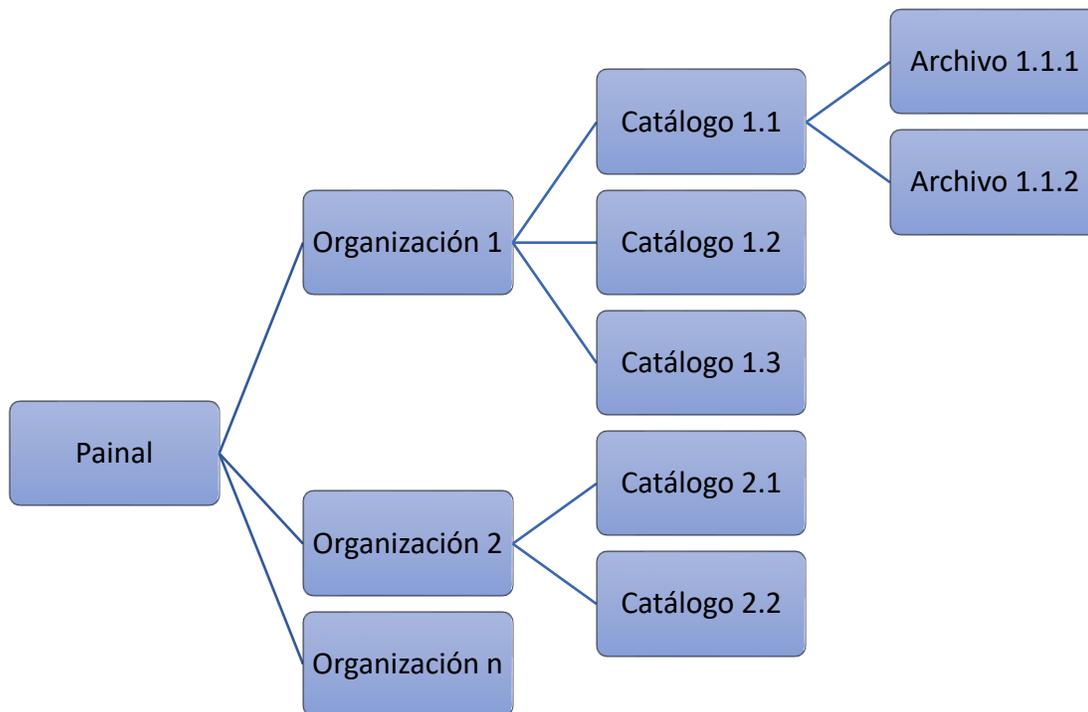


Figura 2 Organización de datos en Painal

## Organizaciones

Entidad independiente que cuenta con su propia jerarquía de usuarios (administradores, publicadores y usuarios finales) que tiene la capacidad de crear catálogos o requiere el acceso a datos de otra organización la cual cuenta con catálogos activos. Las organizaciones pueden contar o no con infraestructura para desplegar un servicio de publicación de catálogos. Sin embargo, pueden acceder a los catálogos creados por otras organizaciones mediante una aplicación cliente, siempre que cuenten con permisos de acceso y los administradores autoricen la suscripción.

## Catálogos

La abstracción de catálogo se refiere al componente utilizado en Painal para la gestión de los metadatos de los contenidos digitales. Un catálogo está diseñado para que las organizaciones puedan tener control sobre los flujos de contenidos generado a través de usuarios (publicadores), consumidores externos que se suscriben a sus contenidos (usuarios finales) y editores que permiten aceptar o rechazar las nuevas publicaciones o suscripciones al contenido de los catálogos (publicadores/administradores).

## Archivos

Contenido digital transportado y almacenado utilizando **Painal**, estos contenidos pueden ser estructurados (hojas de cálculo, bases de datos, documentos de texto, etc.) o no estructurados (señales de audio, imágenes. Tomografías, etc.) El contenido de los archivos compartidos por medio de catálogos solo se encuentra legibles en los catálogos recuperados por usuarios autorizados, en los servicios de almacenamiento solo se encuentran piezas cifradas y dispersas de los contenidos originales, las cuales no pueden ser leídas por terceros o administradores del servicio.

## Alwa

**Alwa** es un repositorio de servicios, datos, contenidos y resultados producidos las comunidades/organizaciones que son parte de los sistemas e-salud. Este servicio de repositorio esta basados en los principios FAIR y es gestionado a través de DSpace, para el almacenamiento y transporte de datos es gestionado por Painal. Los datos y metadatos son depositados y publicados a través de Alwa para poder compartir los datos científicos y tecnológicos obtenidos a través de todos los componentes disponibles de MuyaI.

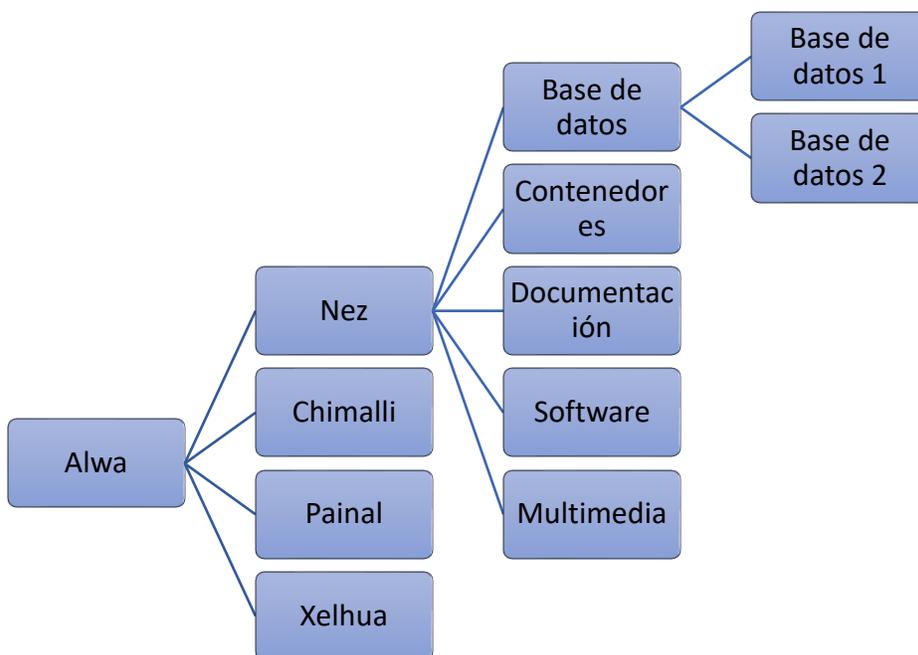


Figura 3 Organización de Alwa con la integración de DSpace y Painal

## Comunidades y organizaciones

Las comunidades y las organizaciones que contiene Alwa son:

- **Nez:** Servicio de construcción de sistemas e-salud.
- **Chimalli:** Servicio de acceso seguro a servicios de e-salud y/o sistemas de analítica.

- **Painal:** Servicio para el transporte y almacenamiento de datos médicos.
- **Xelhua:** Sistema agnóstico en la nube para la construcción de soluciones de big data basada en el diseño de servicios de ciencia de datos de alta disponibilidad y tolerante a fallos.

## Colecciones y catálogos

Las colecciones y catálogos que contiene Alwa son los siguientes:

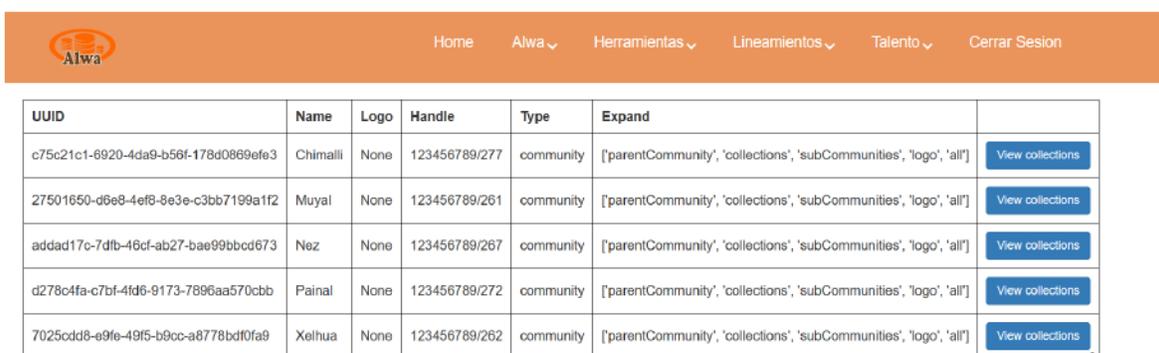
- **Base de datos:** son el conjunto de datos utilizados o generados por el componente de Xelhua.
- **Contenedores (servicios):** son las imágenes de contenedor requeridas para desplegar sistemas de e-salud.
- **Documentación:** es todo aquella documentación, reportes, artículos o manuales de usuarios creado por cada comunidad.
- **Software (código fuente):** es el código fuente de los componentes creados para los sistemas de e-salud.
- **Multimedia:** son todos aquellos archivos que fueron creados para difusión o divulgación de las plataformas (imágenes, videos, audios, etc.).

## Ítems y archivos

Los ítems son los metadatos que son almacenados como fichas técnicas de cada archivo que es almacenado en Painal.

## Implementación del prototipo Alwa

El prototipo del sistema de Alwa se implementó mediante el lenguaje de programación Python, a través del framework Flask. Este servicio se conecta con DSpace para la gestión de los metadatos y con Painal para la gestión de datos.



The screenshot shows the Alwa web interface. At the top, there is an orange navigation bar with the Alwa logo on the left and several menu items: Home, Alwa (with a dropdown arrow), Herramientas (with a dropdown arrow), Lineamientos (with a dropdown arrow), Talento (with a dropdown arrow), and Cerrar Sesión. Below the navigation bar is a table with the following columns: UUID, Name, Logo, Handle, Type, and Expand. Each row represents a community and includes a 'View collections' button.

UUID	Name	Logo	Handle	Type	Expand	
c75c21c1-6920-4da9-b56f-178d0869efe3	Chimalli	None	123456789/277	community	[parentCommunity', 'collections', 'subCommunities', 'logo', 'all']	<a href="#">View collections</a>
27501650-d6e8-4ef8-8e3e-c3bb7199a1f2	Muyal	None	123456789/261	community	[parentCommunity', 'collections', 'subCommunities', 'logo', 'all']	<a href="#">View collections</a>
addad17c-7dfb-46cf-ab27-bae99bbcd673	Nez	None	123456789/267	community	[parentCommunity', 'collections', 'subCommunities', 'logo', 'all']	<a href="#">View collections</a>
d278c4fa-c7bf-4fd6-9173-7896aa570cbb	Painal	None	123456789/272	community	[parentCommunity', 'collections', 'subCommunities', 'logo', 'all']	<a href="#">View collections</a>
7025cdd8-e9fe-49f5-b9cc-a8778bd10fa9	Xelhua	None	123456789/262	community	[parentCommunity', 'collections', 'subCommunities', 'logo', 'all']	<a href="#">View collections</a>

Figura 4 Interfaz gráfica de Alwa con sus respectivas comunidades y organizaciones

Alwa permite a las instituciones visualizar los repositorios de los servicios para crear sistemas de e-salud a través de catálogos o colecciones.

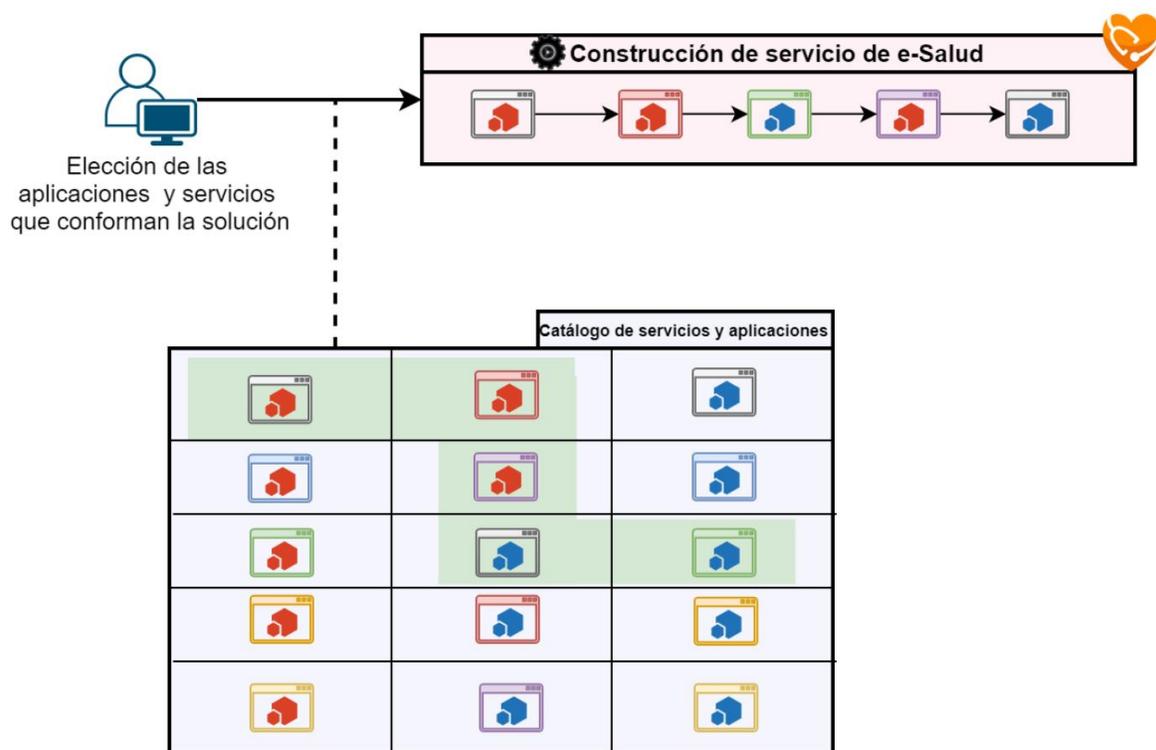


Figura 5 Acceso a los catálogos y colecciones de los servicios disponibles

## Bibliografía

Boeckhout, M. Z. (2018). The FAIR guiding principles for data stewardship: fair enough?. . *European journal of human genetics*, 931-936.

GO FAIR. (n.d.). Retrieved from <https://www.go-fair.org/fair-principles/>

Rodríguez-Gairín, J. M.-D. (2008). *DSpace: un manual específico para gestores de la información y la documentación*. BiD: textos universitaris de biblioteconomia i documentació.