

Universidad Nacional de Lanús - Licenciatura en Sistemas

# Taller de sistemas de gestión editorial y ciencia abierta

## Introducción al software DSpace

14/05/2025

Gonzalo Faramiñan

gonzalo.faraminan@uns.edu.ar

Biblioteca Central “Prof. Nicolás Matijevic”

Universidad Nacional del Sur



Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional

UNIVERSIDAD NACIONAL  
UN  
La.  
DE LANUS

 UNS UNIVERSIDAD  
NACIONAL  
DEL SUR

# Contenidos

- Introducción:
  - Repositorio institucional (RI)
  - software para RI
- DSpace
  - introducción
  - aspectos técnicos relevantes
  - requerimientos/instalación
  - arquitectura
  - administración
  - actualización
  - documentación

# ¿Qué es un Repositorio Institucional (RI)?

Se consideran repositorios digitales a aquellas **colecciones** digitales de la producción científico-tecnológica de una **institución**, en las que se permite la búsqueda y la recuperación de información para su posterior uso nacional e internacional. Un repositorio digital contiene mecanismos para **importar, identificar, almacenar, preservar, recuperar y exportar** un conjunto de **objetos digitales**, normalmente desde un portal web. Esos objetos son **descritos mediante metadatos** que facilitan su recuperación. A su vez, los repositorios digitales son **abiertos e interactivos**, pues cumplen con **protocolos** internacionales que permiten la **interoperabilidad** entre ellos.

Fuente: Sistema Nacional de Repositorios Digitales (2015). *Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales*.

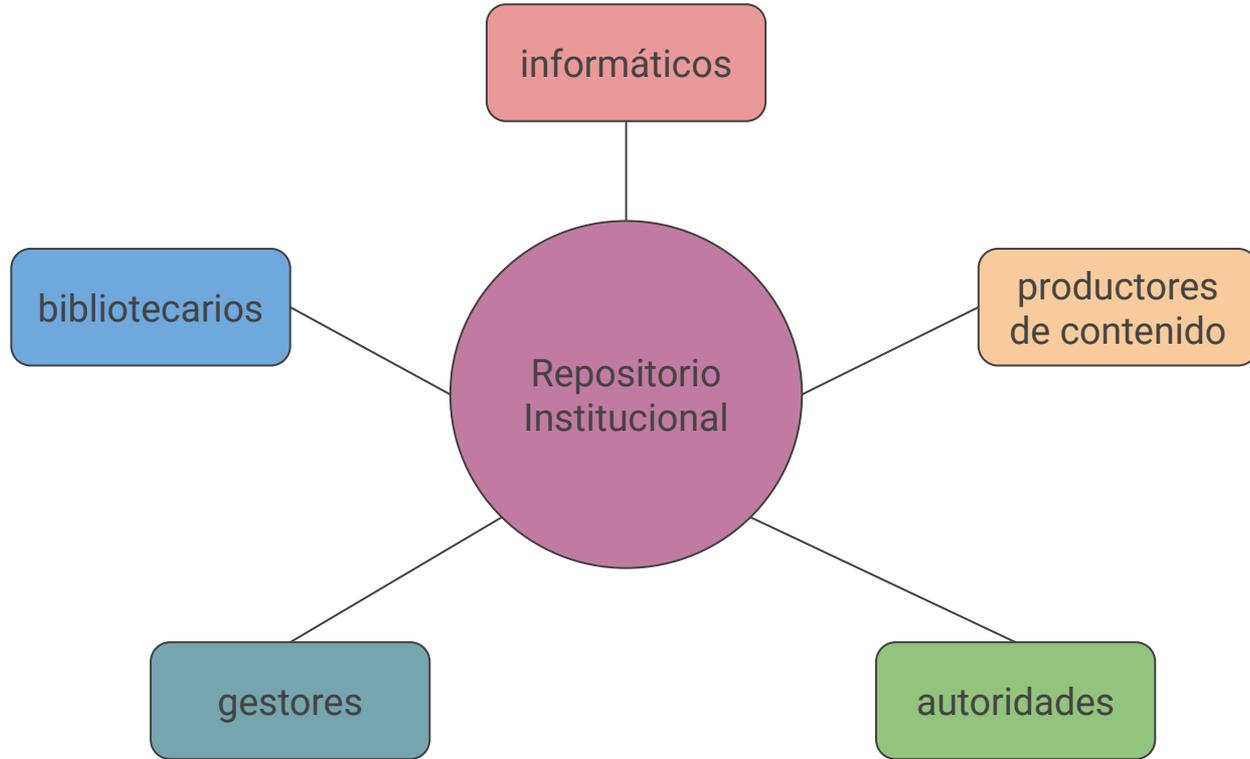
# Características generales de un software de RI

- **Gestión de objetos digitales:** carga de documentos, organización en colecciones, versionado de ítems, importación/exportación
- **Gestión de metadatos:** descripción estructurada de OD, soporte para múltiples esquemas (Dublin Core, MODS, LOM, etc.), importación/exportación, mapeos para directrices
- **Búsqueda y descubrimiento:** catálogo con índices (autor, palabra clave, etc), navegación (*browsing*), facetas o filtros
- **Control de acceso y permisos:** gestión de usuarios y grupos, integración con servicios de autenticación (p.e. LDAP)
- **Flujos de trabajo:** workflows personalizables para la gestión de OD y metadatos

# Características generales de un software de RI

- **Interoperabilidad:** OAI-PMH para cosecha por agregadores, integración para proveedores de servicios (ORCID, DOI, etc)
- **Estadísticas:** de acceso y descarga, integración con herramientas externas (p.e. Google Analytics, Matomo)
- **Preservación digital:** soporte de checksums, soporte para múltiples formatos/archivos por ítem
- **Configuración y personalización:** panel administrativo, personalización en un espacio local que interactúe con el core, branding para frontend, integración por API

# Roles involucrados en un proyecto de RI



# Opciones open source para RI



greenstone digital library software

## FedoraCommons

88% de los repositorios de la región utiliza DSpace\*

# DSpace: introducción

## - Gestión del proyecto:

- 2002: primera versión (MIT y HP Labs)
- 2007: DSpace Foundation (fusión de MIT y HP)
- 2009: DuraSpace (fusión de DS Foundation y Fedora Commons)
- 2019: DSpace Program (DSpace Governance, fusión de DuraDSpace y Lyrisis).

Liderazgo de Lyrisis



- **Licencia:** BSD License (Berkeley Software Distribution); libre

## - Enlaces:

- <https://dspace.org/>
- <https://github.com/DSpace/DSpace>

# Aspectos técnicos relevantes de DSpace (v7+)

- **Interfaz web (UI):**
  - accesibilidad web WCAG 2.1 AA (accesibilidad intermedia)
  - altamente personalizable
- **Rest API:** completa para todas las acciones del backend
  - en DS 7 se puede usar también la API de DS 6
- **Seguridad:**
  - uso forzoso de HTTPS en producción
  - CORS para establecer dominios de confianza que consumen la API
- **Actualización:**
  - backend: directa a última estable desde v1.x
- **Documentación** amplia, **comunidad** activa y predispuesta



# Requerimientos: backend

- **Framework:** Spring Boot



- **Ecosistema Java:**

- Java JDK (OpenJDK u Oracle JDK)

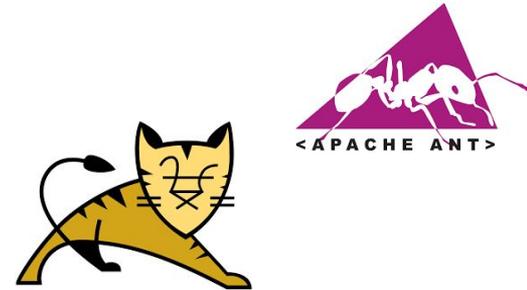
- **Build:**

- Apache Maven; empaquetado, armado del instalador

- Apache Ant; deploy al directorio de instalación

- Servidor de aplicaciones (servlet container): Apache Tomcat

**Maven™**



- **Base de datos:** PostgreSQL



PostgreSQL

- **Motor de búsqueda:** Apache Solr

Solr

# Requerimientos: frontend

- **Framework:** Angular

- **Ecosistema JavaScript:**

  - Build:

    - Node.js; también servidor

    - Yarn

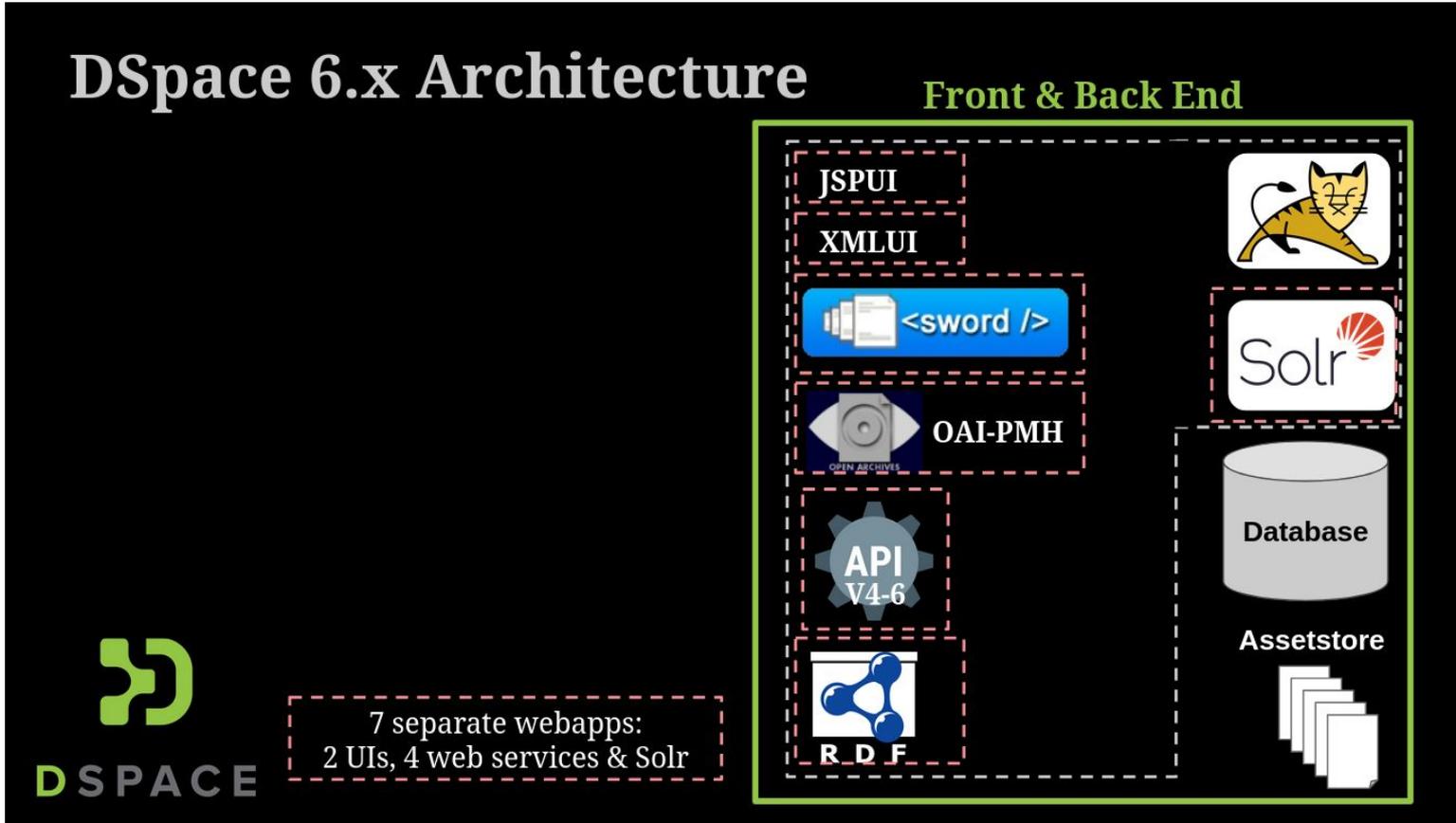
- **Administrador de procesos:** PM2

→ para completar la instalación necesitamos un

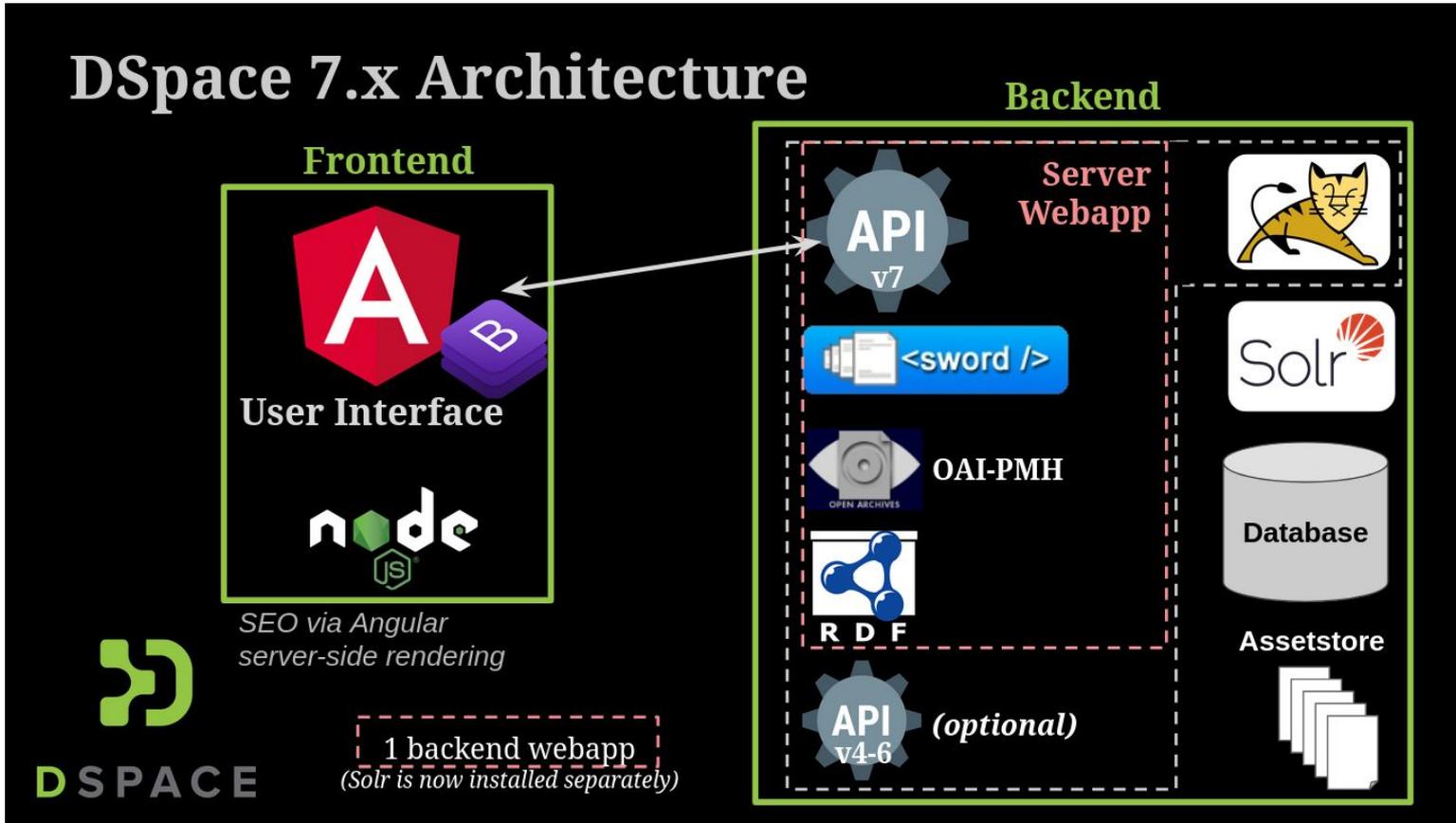
**servidor web** como proxy reverso (Apache HTTPD o Nginx)



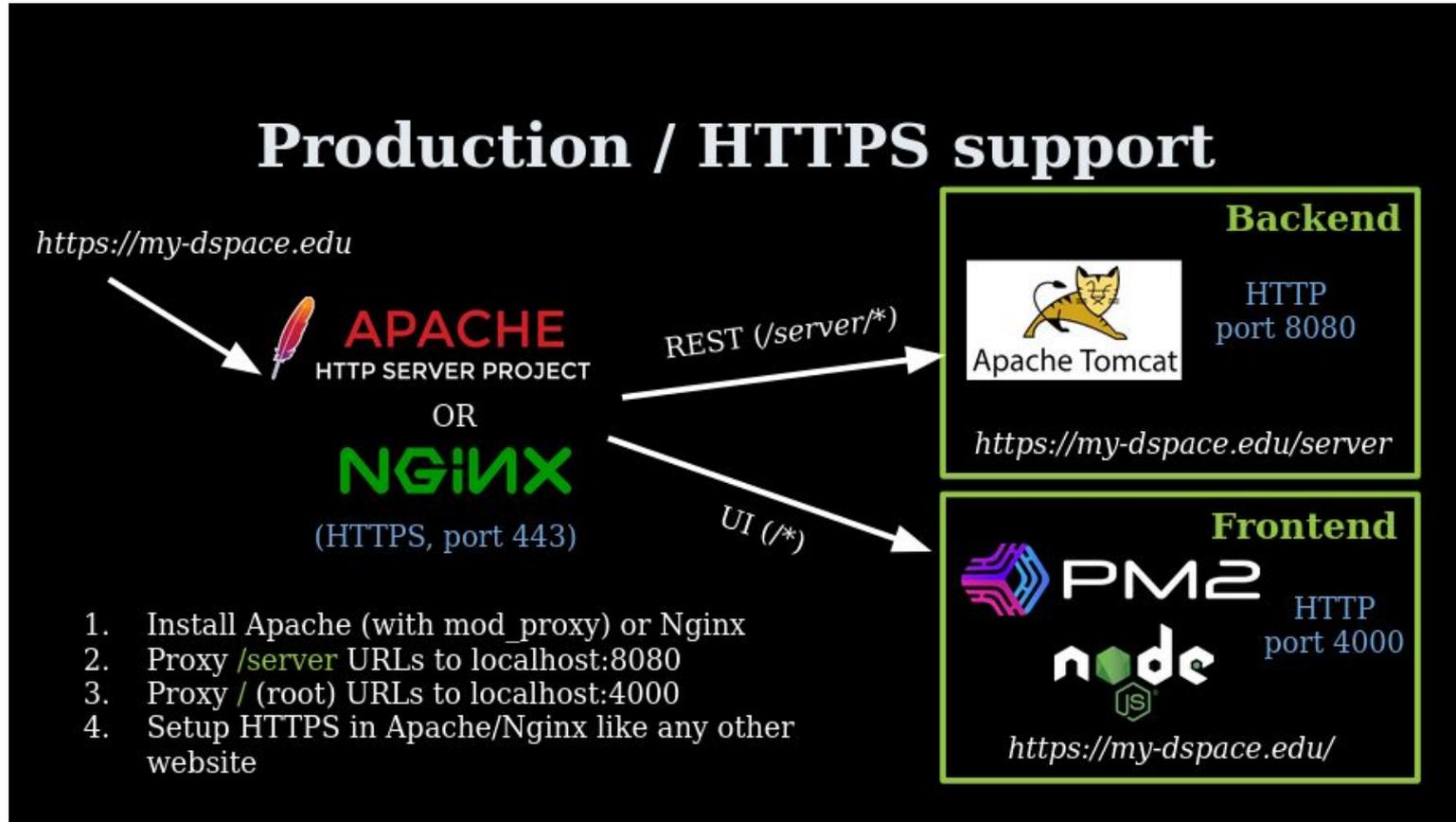
# DSpace 6.x: arquitectura



# DSpace 7.x: arquitectura



# DSpace 7.x: arquitectura



# Administración de DSpace

## a. Administración web

- limitación en tareas disponibles
- se están incorporando más acciones en las últimas versiones

Home • Health

### Health

Status Info

Status : 

Database	 
Geoip	 
Solr: Authority Core	 
Solr: Oai Core	 
Solr: Search Core	 
Solr: Statistics Core	 

# Administración de DSpace

## b. Administración desde el servidor/SSH

- workflows, formularios de carga, listados de vocabularios controlados, etc.
- opciones de configuración y herramientas de administración para:
  - afrontar desafíos de performance (SSR, memoria, Solr, etc)
  - configuración local
  - mapear y transformar metadatos para OAI-PMH
  - SEO
  - etc.

# Administración de DSpace

## b. Administración desde el servidor/SSH

- utilitario para CLI `/bin/dspace`
  - programar de tareas en el server (`cron`)
  - ejecutar tareas a demanda
    - - manipulación de bitstreams (`assetstore`)
    - generación de miniaturas e indexación a texto completo (`filter-media`)
    - mantenimiento de los índices de Solr
    - gestión del servidor OAI-PMH (`service provider`)
    - exportación/importación/edición de metadatos y OD en lote
    - gestión de notificaciones a usuarios
    - etc

# Administración de DSpace

## b. Administración desde el servidor/SSH

- utilitario para CLI `/bin/dspace`

```
/bin/dspace filter-media -p "PDF Text Extractor" -i 123456789/3422
```

→ extraer texto (nativo/OCR) de archivo(s) PDF para indexación a texto completo de los bitstream contenidos en el handle indicado (comunidad/colección/item)

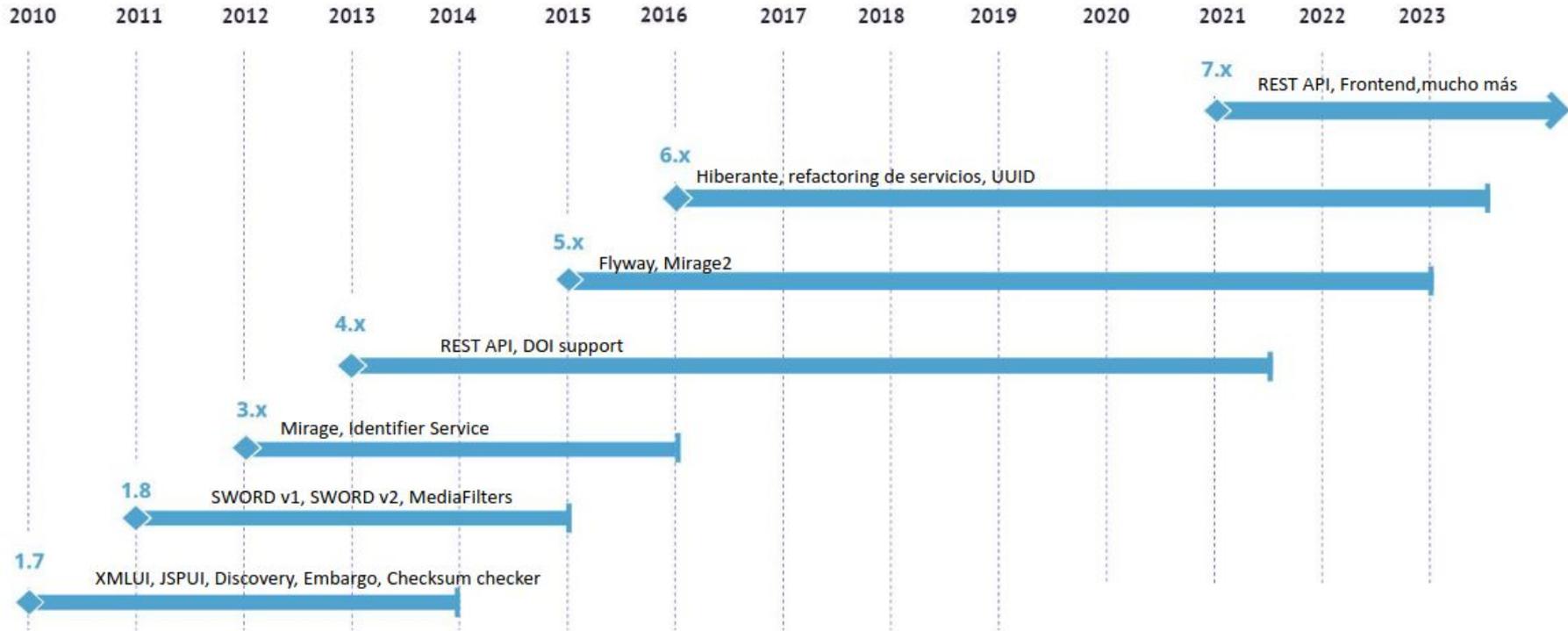
# Administración de DSpace

## b. Administración desde el servidor/SSH

- programación de tareas en el server (**cron**)

```
#-----  
# GLOBAL VARIABLES  
#-----  
DSPACE = /var/www/dspace  
SHELL=/bin/sh  
PATH=/usr/local/sbin:/usr/local/bin:/sbin:/bin:/usr/sbin:/usr/bin  
JAVA_OPTS="-Xmx512M -Xms512M -Dfile.encoding=UTF-8"  
  
#-----  
# DAILY TASKS  
#-----  
  
# Update the OAI-PMH index with the newest content at midnight every day  
0 0 * * * $DSPACE/bin/dspace oai import > /dev/null  
  
# Clean and Update the Discovery indexes at midnight every day  
0 0 * * * $DSPACE/bin/dspace index-discovery > /dev/null  
  
# run the index-authority script to ensure the Solr Authority cache is up to date  
45 0 * * * $DSPACE/bin/dspace index-authority > /dev/null  
  
# Cleanup Web Spiders from DSpace Statistics Solr Index at 01:00 every day  
0 1 * * * $DSPACE/bin/dspace stats-util -f  
  
# Send out "daily" update subscription e-mails to any users who have "subscribed" to a Community/Collection  
0 2 * * * $DSPACE/bin/dspace subscription-send -f D  
  
# Run the media filter (thumbnails and full text indexing)  
0 3 * * * $DSPACE/bin/dspace filter-media  
  
#-----  
# MONTHLY TASKS  
#-----  
  
# Send out "monthly" update subscription  
0 2 1 * * $DSPACE/bin/dspace subscription-send -f M  
  
# Permanently delete any bitstreams flagged as "deleted" in DSpace  
# REQUIRED to fully remove deleted content files from the "assetstore" folder  
0 1 1 * * $DSPACE/bin/dspace cleanup > /dev/null  
..
```

# DSpace: evolución de versiones



Fuente: Lira, de Albuquerque, de Giusti (2023). *Experiencias en la migración de repositorios institucionales DSpace*

# Actualización de DSpace: política de soporte

Soporte, *bug fixes* y actualizaciones de seguridad para las últimas tres versiones principales

Versiones:

- 9.0 (04/2025) → testathon del release candidate
- 8.0 (06/2024) → *última versión estable actual*
- 7.0 (07/2021) → con soporte
- 6.0 (10/2016) → EOL anticipado en 07/2023
- 5.0 (01/2015) → EOL anticipado en 01/2023

# Actualización de DSpace

## - Planificar el proceso:

- ¿hay personalizaciones? ¿de qué tipo? ¿en front, back? ¿hemos realizado buenas prácticas (VCS, local)?
- versión destino: release notes, requerimientos
  - ¿se puede instalar en entorno de pruebas?
- evaluar/gestionar recursos (personal, infraestructura, escalabilidad, etc)
- plazos: estimar, acordar con autoridades y comunicar a las partes interesadas
- establecer una hoja de ruta
- documentar

## - **Documentación** amplia, **comunidad** activa y predispuesta

# Actualización de DSpace

*Se puede actualizar de cualquier versión de DSpace (1.x, 3.x, 4.x, 5.x, 6.x, 7.x, 8.x) a la última versión de DSpace 8.x (última estable), sin versiones intermedias*



Si actualizamos desde  $\leq 6.x$

- solo se puede actualizar el backend; perdemos el front
- oportunidad para **migrar** a un nuevo server

## a. actualización en el mismo server

1. actualización de requerimientos
2. actualización de backend (base de datos y assetstore)
3. actualización ( $\geq v7.x$ ) o instalación ( $\leq v6.x$ ) del front

# Actualización de DSpace

## b. actualización en server nuevo

1. instalación de requerimientos
2. instalación de backend
3. instalación de frontend
4. migración de datos
  - 4.a. migración de base de datos: eliminar base nueva, restore del dump

```
./dspace database clean
```

### 4.b. migración del assetstore

4.c. actualizar base: `./dspace database migrate ignored`

4.d. reindizar `./dspace index-discovery -b`

```
./dspace oai import
```

# Enlaces de interés

- [https://dspace.org/wp-content/uploads/2022/01/DSpace\\_Diagram.pdf](https://dspace.org/wp-content/uploads/2022/01/DSpace_Diagram.pdf)
- DSpace
  - documentation: <https://wiki.lyrasis.org/display/DSPACE/Home>
  - Wiki: <https://wiki.lyrasis.org/display/DSPACE/Home>
  - v8.x: <https://wiki.lyrasis.org/display/DSDOC8x/Dspace+8.x+Documentation>
- Demo: <https://demo.dspace.org/>
- Comunidad (Slack, mailing): <https://wiki.lyrasis.org/display/DSPACE/Support>
- GitHub
  - UI: <https://github.com/DSpace/dspace-angular/>
  - Rest API: <https://github.com/DSpace/DSpace>



**¡Gracias!**

**¿Preguntas?**