



Experiencias en la migración de repositorios institucionales DSpace

Lira, Ariel J.*; de Albuquerque, Pablo C. *; De Giusti, Marisa R. *

* PREBI-SEDICI, Universidad Nacional de La Plata;
CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Taller: Instalación, configuración básica y experiencias sobre DSpace7
Viernes 19 de Mayo de 2023



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)




EDUCACIÓN
PÚBLICA
Y GRATUITA



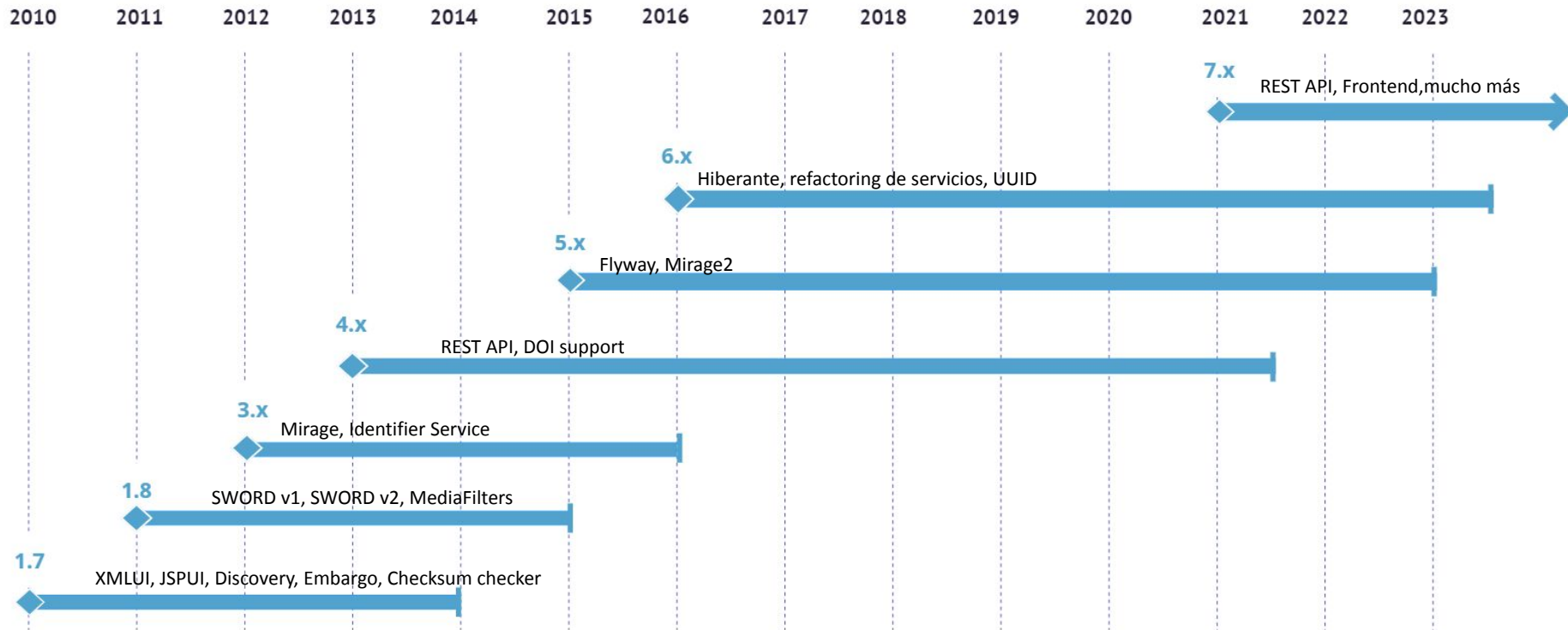
UNIVERSIDAD
NACIONAL
DE LA PLATA



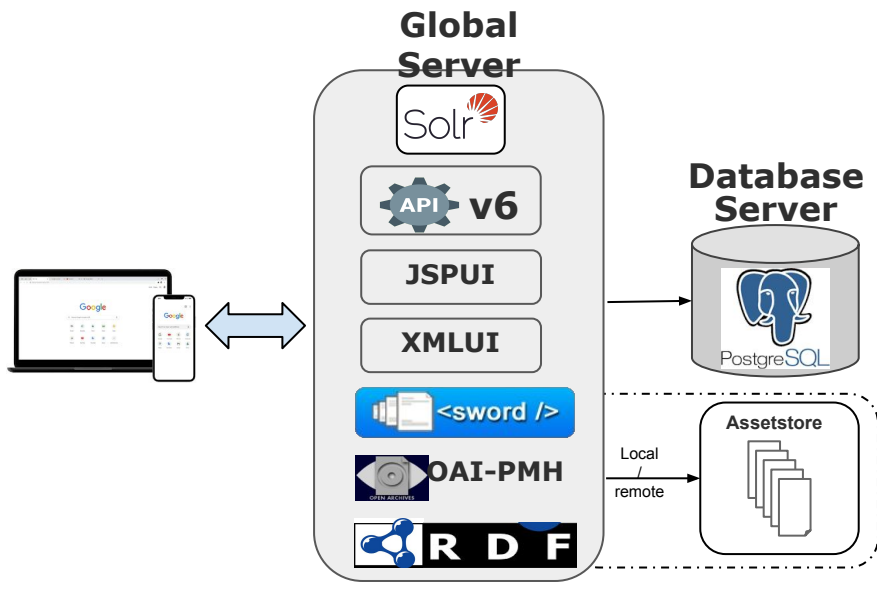
Temario

- Presentar diferencias entre versiones del software DSpace y analizar su adopción por parte de la comunidad
 - Relatar 2 experiencias de migración
 - CIC-DIGITAL <https://digital.cic.gba.gob.ar/>
 - SEDICI <http://sedici.unlp.edu.ar/>
 - Plantear caminos de migración para al personal técnico encargado de aplicar los cambios
 - Resumir aspectos a tener en cuenta para tomar la decisión de migrar (o no)
- 

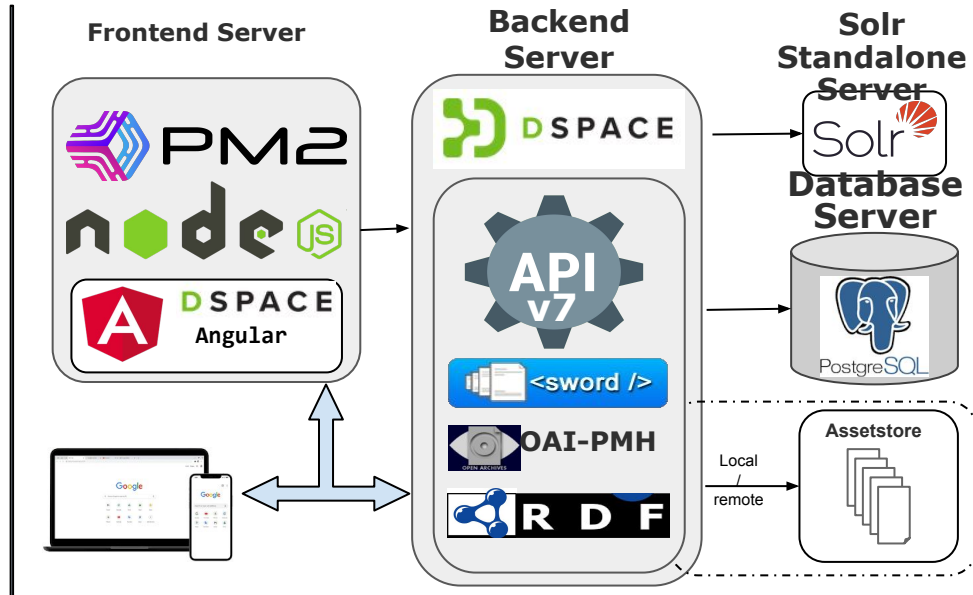
Evolución de versiones



Cambios de Arquitectura



DSpace 6 y anteriores

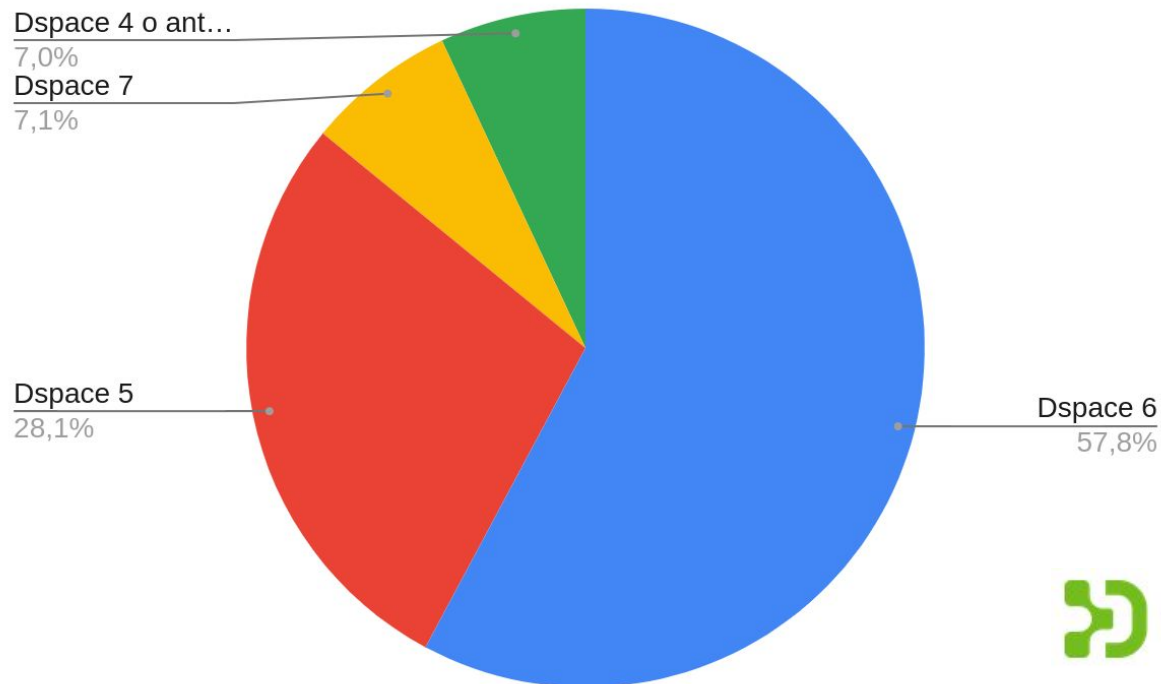


DSpace 7

Funcionalidad incorporada en **DSpace 7**

- Interfaz más amigable con el usuario
- Mejoras varias en SEO, WCAG 2.1
- Seguridad: tecnologías actuales, uso forzado de https y CORS
- Nuevo Backend con api rest estándar
- Posibilidad de extender el modelo con entidades CRIS
 - Personas, Organizaciones, Proyectos,
- Posibilidad de extender el modelo con otras entidades
 - Revistas, volúmenes y números,
 - + Eventos, ediciones, etc,
- y mucha funcionalidad más:
 - importación automática de metadatos a partir de múltiples servicios
 - dashboard de admin con buscador y facets
 - integración con ORCID

Uso de versiones **DSpace** en todo el mundo



Fuente: análisis propio a partir del tag `<meta name="Generator" />` de 1107 repositorios considerados entre un total de 2420 repositorios. Listado extraído de opendoar con filtro "software DSpace" <http://roar.eprints.org/>

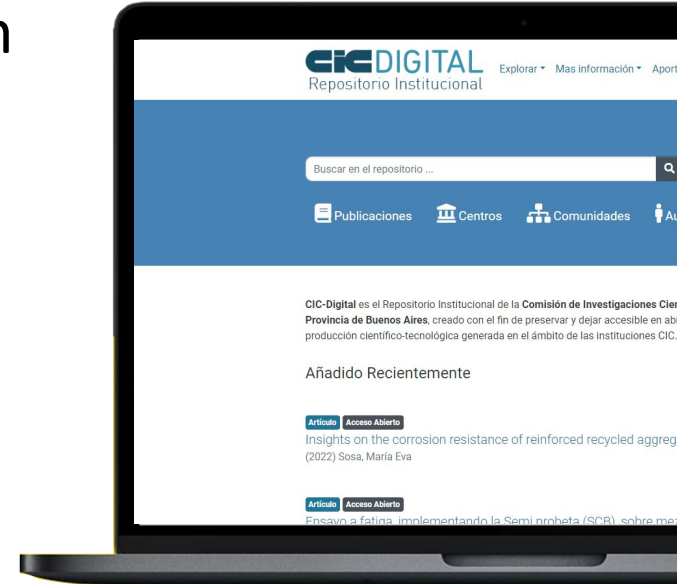
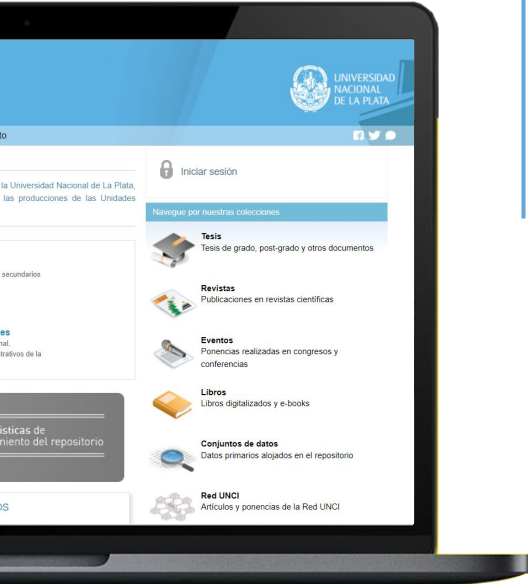
Bajo nivel de actualización

- Causas posibles
 - Falta de personal informáticos dedicado propio
 - Falta de soporte desde las áreas de sistemas de las instituciones
 - Senda de migración poco clara para quienes no se dedican a mantener este tipo de software y compleja
- Problemas
 - Baja calidad de servicio
 - Versiones sin soporte (EOL)
 - Versiones inseguras

Bajo nivel de actualización - ¿Cómo mejorar?

- Estrategias:
 - Conseguir personal informático abocado al repositorio
 - Compromiso institucional
 - Cooperación en red
 - con financiamiento desde miembros y/o de los organismos nacionales
 - hay muchos casos de éxito

Experiencia de migración en 2 repositorios institucionales



Escenario

SEDICI

SERVICIO DE DIFUSIÓN
DE LA CREACIÓN INTELECTUAL



<http://sedici.unlp.edu.ar/>

Repositorio institucional de la
Universidad Nacional de La Plata, Argentina

Año creación 2003

Total publicaciones: **144.000**

Crecimiento prom. mensual (9 años): **1000**

Crecimiento 2022: **17.000**

Escenario

CIC DIGITAL

Repositorio Institucional

<https://digital.cic.gba.gob.ar>

Repositorio institucional de la **Comisión de
Investigaciones Científicas, Argentina**

Año creación 2014

Total publicaciones: **9200**

Crecimiento prom. mensual (9 años): **100**

Crecimiento 2022: **300**

Estadísticas de uso CIC-DIGITAL vs SEDICI



Estadísticas de uso

Sesiones

3.600.537

Usuarios

2.699.518

250.000 requests x día

60% son bots buenos

40% usuarios + bots malos

Estadísticas de uso

Sesiones

1.778.943

Usuarios

1.760.514

<50.000 requests x día

60-80% son bots buenos

resto usuarios + bots malos

Nivel de Personalización

SEDICI

- **1500** commits y **12** committers en **12** años
- <https://github.com/sedici/DSpace>
- Muchas personalizaciones
 - Workflow y submission por comunidades,
 - extensiones al submission con negaciones de type-bind y roles,
 - edición de ítems publicados desde workflow,
 - Mejoras para performance
 - Vocabularios controlados
 - Open Search

CIC-DIGITAL

- **800** commits y **5** committers en **7** años
- <https://github.com/cicba/DSpace>
- <https://github.com/cicba/DSpace-angular>
- Pocas personalizaciones
 - Request a Copy
 - Vocabularios controlados

Roadmap **CiC**DIGITAL

Repositorio Institucional

DSPACE 7.0 Testathon

- Participación en Testathon
- Merge en desarrollo
- Prototipo de UI basado en Mantis

Merge con **DSPACE 7.1**

- Merge en desarrollo
- Nueva UI basada en theme "custom"

Merge con **DSPACE 7.2.1**

- Pasamos a producción..
- Merge con 7.2 y después con 7.2.1



Versión de
DSpace

CIC 6



Repositorio Institucional
Comisión de Investigaciones Científicas

[Inicio](#) [Explorar](#) [Aportar Material](#) [Mas información](#) [Contacto](#)

[Mi cuenta](#) [Estadísticas](#) [ES](#)

CIC-Digital es el Repositorio Institucional de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, creado con el fin de preservar y dejar accesible en abierto toda la producción científico-tecnológica generada en el ámbito de las instituciones CIC.

BUSCAR EN EL REPOSITORIO (10462 ITEMS DISPONIBLES)

[EXPLORAR AUTORES](#) | [EXPLORAR CENTROS](#) | [TODO EL REPOSITORIO](#)

Añadido Recientemente

Reportes estadísticos para repositorios digitales a partir de múltiples fuentes basado en el stack ELK
de Abujerque, Pablo César, Villarreal, Gonzalo Luján; De Gusti, Marisa Raquel

Relevamiento de repositorios con perfiles de autor
Manzur, Ezequiel; Testamanti, Santiago; Villarreal, Gonzalo Luján; De Gusti, Marisa Raquel; Lira, Ariel Jorge

Acta n° 1540
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Acta n° 1539
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Acta n° 1359
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Acta n° 1358
Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Subir Material

A través del proceso de **autoarchivo** usted podrá contribuir con sus investigaciones y desarrollos científicos propios, al contenido dentro del Repositorio Institucional de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires.

Más información

CIC 7



Repositorio Institucional

[Explorar](#) [Mas información](#) [Aportar material](#)

[Iniciar sesión](#)

[Buscar](#)

[Publicaciones](#) [Centros](#) [Comunidades](#) [Autores](#)



CIC-Digital es el Repositorio Institucional de la **Comisión de Investigaciones Científicas de la Provincia de Buenos Aires**, creado con el fin de preservar y dejar accesible en abierto toda la producción científico-tecnológica generada en el ámbito de las instituciones CIC. [Leer más](#)

Añadido Recientemente

[Actas de directorio](#) [Acceso Abierto](#)

Acta n° 1563
(2022) Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

[Documento de conferencia](#) [Acceso Abierto](#)

[La dimensión ambiental en la formación tecnológica de los diseñadores industriales de la Universidad Nacional de La Plata](#)

(2022) Del Giorgio Solfa, Federico; Alfano, Alejandro César

Subir material

Si sos miembro de la **Comisión de Investigaciones Científicas (CIC)** o trabajás en alguno de sus centros o centros asociados, te invitamos a depositar tus publicaciones y datos de investigación.


¿Por qué usar CIC-DIGITAL?

- **Impacto** - La visibilidad que otorga el repositorio permite una gran difusión y reuso de las obras, lo que resulta en mayor impacto de la producción para los autores y la institución.
- **Difusión rápida y efectiva** - El repositorio expone las obras de

Problemas encontrados

- URL de ítems basadas en uuid en lugar de handle
 - Antes: <https://digital.cic.gba.gob.ar/handle/11746/11875>
 - Despues <https://digital.cic.gba.gob.ar/items/61a3f173-5a93-4439-9323-ff8b9541ac67>
 - Se hace redirección del handle al uuid
 - Posible impacto negativo en visibilidad
 - Confusión para el usuario
- Google Scholar referencia a las urls del backend al indexar los bitstreams
 - Posible impacto negativo en visibilidad o rankings (ej webometrics) si el dominio es diferente al de frontend
 - <https://digital.cic.gba.gob.ar/bitstreams/9ff250ab-8014-4805-91e8-d48259c320ca/download> (302)
 - <https://host170.sedici.unlp.edu.ar/server/api/core/bitstreams/9ff250ab-8014-4805-91e8-d48259c320ca/content>

Problemas encontrados

- Falta de customización de algunos componentes
 - se pueden personalizar, aunque posteriormente puede dar conflictos en próximas versiones cuando se hagan personalizables
 - No se puede cambiar colección de un ítem mientras está en workflow
 - No se puede modificar metadatos controlados por autoridad desde la admin
 - Cuando se tienen comunidades con el mismo nombre, en algunos componentes se dificulta diferenciar los mismos.
- 

Roadmap



Experiencias de migración en **SEDICI**

Prueba de migración con código a DSpace 6

- performance con hibernate
- muchos conflictos por el el refactoring de servicios que se hizo de 5 a 6

Prueba de migración sin código a DSpace 7 (en abril de 2020)

- Problemas de performance, sumado a la cantidad de recursos y comunidades.
- Workflow de SEDICI muy personalizado

SEDICI. Plan de migración 2023-2024

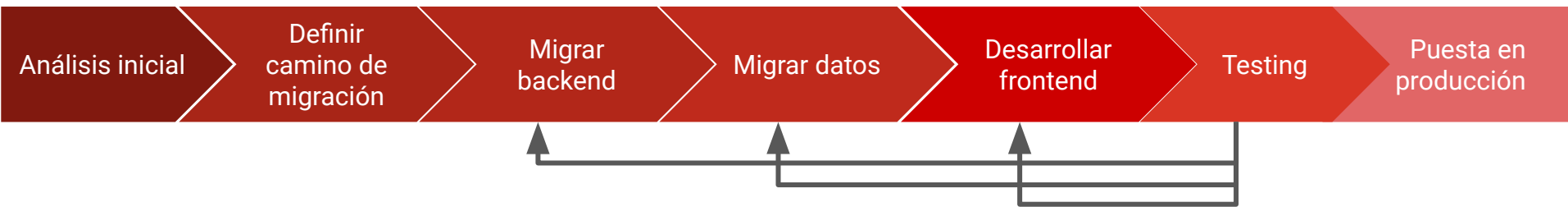
Por la cantidad de cambios entre SEDICI/DSpace (5.10) y DSpace/DSpace (7.x) es casi imposible hacer un merge con 7.x.

Hay un plan:

- Instalar DSpace desde cero
- Re-implementar/ajustar funcionalidad
 - edición desde submission
 - configuración de submission y workflow en comunidades
 - garantizar niveles de performance
 - garantizar indexación correcta en scholar y google

Será un proceso complejo y largo para llegar a DSpace 7 u 8.

Estrategias de migración



Revisar lo que se tiene

- Relevar personalizaciones hechas
 - en backend
 - en frontend
- Compatibilidad de features entre lo que tenemos y lo que ofrece DSpace 7.x
 - ver [What features are coming in a later 7.x release?](#)

Priorización de requerimientos

- personalizaciones a descartar
- features a implementar
- definir requerimientos del nuevo diseño

Tier 2: Medium-High Priority

See tasks on our [Tier 2 Board](#)

1. ✓ (Available in 7.3) (NEW) Advance
2. ✓ (Available in 7.3) (NEW) Migrate
3. ✓ (Available in 7.5) (General) Supp
not in Search. (UI ticket #815, RES
4. ✓ (Available in 7.4) (General) List c
5. ✓ (Available in 7.3) (Submission) SI
6. ⚠ (Delayed, requires further analy
7. ⚠ (Delayed, as this feature is highl
submission. In 7.x, metadata extra
8. ✓ (Available in 7.5) (Submission) Sv
9. ✓ (Available in 7.2) (General) Supp
configuration. (UI ticket #852)
10. ✓ (Available in 7.3) (Admin) Admin
11. ✓ (Available in 7.3 - donated by Th
12. ✓ (Available in 7.2) (Submission) M
13. ⓘ (Under development. Likely in 7.
14. ✓ (Available in 7.4) (Integrations) C
15. ✓ (Available in 7.3) (NEW) Version
16. ✓ (Available in 7.5) (Authenticatio

Análisis inicial

Definir
camino de
migración

Migrar
backend

Migrar
datos

Desarrollar
frontend

Testing

Puesta en
producción

Opción a) Migrar personalizaciones (de backend) vía merge de código

- Mantiene mejoras, fixes, módulos, etc
- Mantiene historial de cambios interno

Válido para:

- repositorios con personalización relevante Y con versionado en un repositorio git

Análisis
inicial

Definir
camino de
migración

Migrar
backend

Migrar
datos

Desarrollar
frontend

Testing

Puesta en
producción

Opción b) Hacer una instalación nueva de DSpace 7

- Desarrollo más prolijo iniciado de cero
- Se pierde el historial de trabajo y personalizaciones
- Puede requerir reimplementar funcionalidad

Válido para:

- Repositorios existentes con muy baja personalización
 - El costo de replicar personalización es bajo
- Repositorios con personalización relevante pero SIN versionado de código
 - Es el único camino.

Análisis inicial

Definir camino de migración

Migrar backend

Migrar datos

Desarrollar frontend

Testing

Puesta en producción

- Migrar código de backend

- git-merge con el código de DSpace/Dspace
- Resolución de conflictos
 - considerar backports, hotfixes aplicados
- Adaptar personalizaciones
 - Conectores con autoridades
 - extensiones al modelo o base de datos
 - cambios en “additions”

- Migrar configuraciones

- Mapeos (OAI, scholar)
- Actualizar Workflow y formularios de carga
- Revisar configs removidas o cambiadas (log4j, workflow, input-forms, etc)
- ver documento de migración ds DSpace

<https://wiki.lyrasis.org/display/DSDOC7x/Upgrading+DSpace>

Aquí es vital contar con cambios documentados a partir de tickets, commits, mails y/o documentación interna.

Análisis inicial

**Definir
camino de
migración**

**Migrar
backend**

**Migrar
datos**

Desarrollar
frontend

Testing

Puesta en
producción

- Base de datos
 - Flyway Migrations
- Solr cores
 - Authority
 - Statistics
- Assetstore

Análisis inicial

**Definir
camino de
migración**

**Migrar
backend**

**Migrar
datos**

**Desarrollar
frontend**

Testing

Puesta en
producción

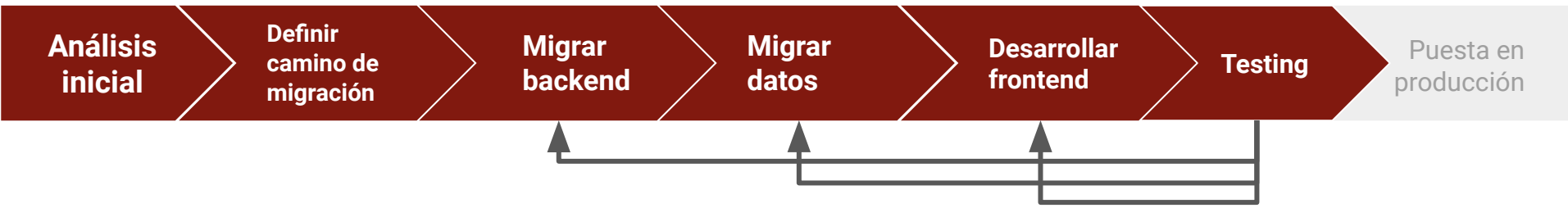
Armado de mockups al menos de

- página de inicio,
- vista de un ítem
- página de búsqueda

Minimizar personalización de diseño, al menos al principio

Implementación en DSpace-angular

Validar vista responsive (al menos un 25% del público usa móviles)



- Pruebas en un servidor de testing por catalogadores
 - hay que estar seguros que lo que la nueva versión provee es lo que se precisa y acepta.
- Pruebas con usuarios externos/públicos
- Pruebas en múltiples contextos: celulares, conexiones lentas, bloqueo de cookies, acceso desde bots, etc

Análisis inicial

Definir camino de migración

Migrar backend

Migrar datos

Desarrollar frontend

Testing


Puesta en producción

- Utilizar pm2 en frontend y hacer pruebas de carga
 - ej artillery.io u otras (load, stress)
- usar proxy reverso delante de frontend y backend
 - más seguridad
 - más flexibilidad
- Separar servicios en distintas vms según nivel de carga
 - Ej vms para: frontend, backend, solr, postgres, reverso.
- Importante: conviene usar servidores virtuales nuevos con versiones de SO actuales y en lo posible de líneas LTS

¿Entonces?




¿Por qué no migrar?

- Stack de tecnologías nuevo → mayor curva de aprendizaje
 - Más complejidad: backend java + frontend angular
 - Costo alto
 - ~~Funcionalidad de DSpace6 faltante~~
 - ~~DSpace 7 aún está en activo desarrollo y cambia mucho entre releases~~
- 


¿Por qué migrar?

A corto plazo

- Mejora en la experiencia de usuario
 - Actualización tecnológica
 - Mejoras para SEO
 - Seguridad
 - más personal capacitado con conocimientos
 - Nuevas funcionalidades
- 

¿Por qué migrar?

A mediano plazo

- Entidades configurables para armar Perfiles de autor, Unidades de la institución, Proyectos, etc
 - Desacoplamiento del backend y posibilidad de agregar nuevas apps “clientes” en el futuro
 - Próximamente:
 - Alineación con Next Generation Repositories (COAR)
 - Signposting
 - ResourceSync
 - Notify Project
- 

Es necesario hacer un análisis de **costo-beneficio**

1. Desarrollar un plan de migración
 - a. Análisis técnico para definir la complejidad del proceso
 - b. Relevar cantidad y complejidad de personalizaciones
 - c. Definir senda de migración
 - d. Considerar capacitación del grupo de trabajo
 - e. Estimar plazo necesario para migrar
 - f. Estimación de costo (recursos humanos y económicos)
 - g. Evaluación de beneficios de migrar y riesgos de no migrar
2. Lograr decisión y compromiso institucional
3. Comenzar la migración de acuerdo al plan

Referencias

- [What features are coming in a later 7.x release?](#) (Dspace Wiki)
- Upgrading DSpace
<https://wiki.lyrasis.org/display/DSDOC7x/Upgrading+DSpace> (Dspace Wiki)
- Upgrading to DSpace 7.4. Tim Donohue (LYRASIS) and Corrado Lombardi (4Science) . Webinar November 16, 2022.
<https://www.youtube.com/watch?v=2ECza9Zd5PM>
- Experiencias en la migración de un repositorio institucional a DSpace 7; Actas de la XI Conferencia Internacional de Bibliotecas y Repositorios Digitales <http://sedici.unlp.edu.ar/handle/10915/145741>

¡Muchas gracias!

¿Preguntas?

{alira, pablo, marisa.degiusti}@sedici.unlp.edu.ar
CIC-DIGITAL <https://digital.cic.gba.gob.ar/>
SEDICI <http://sedici.unlp.edu.ar/>

