

PASIG 2019

Cuando la preservación digital se convierte en un trabajo común en la BnF

When digital preservation at BnF becomes business as usual...

**Biblioteca nacional de Francia
Bibliothèque nationale de France**

Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

La Biblioteca nacional de Francia



Una institución publica autónoma

2 200 agentes et docenas de profesiones, un presupuesto de ~230 M€

Misiones local, nacional y internacional

Alrededor de 1 millón de lectores por año, 300 000 visitantes a las exposiciones.

Importantes colecciones de gestionar

Más de 14M libros

Más de 30M carteles y fotos

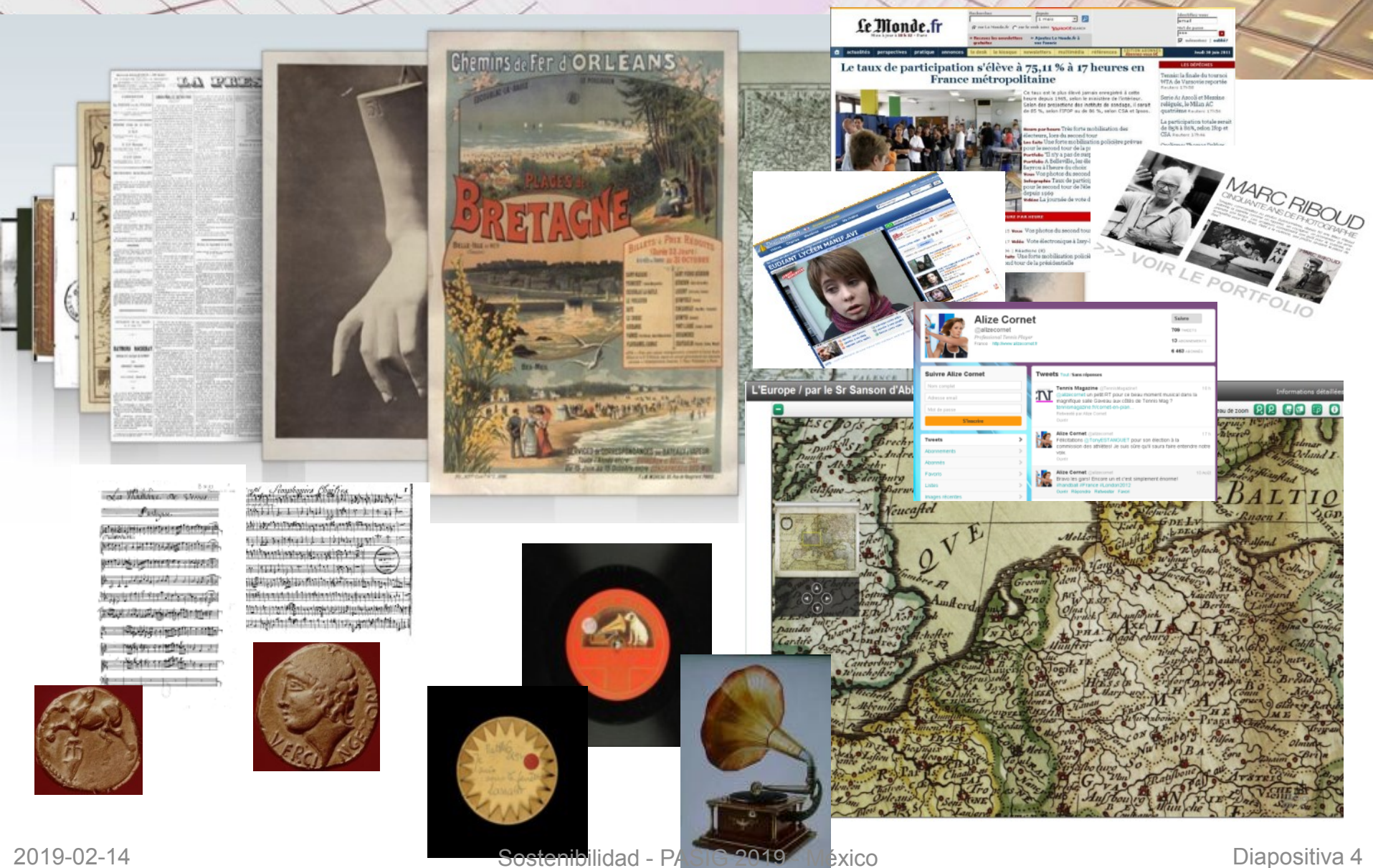
Más de 250.000 manuscritos

Más de 1M de material audiovisual

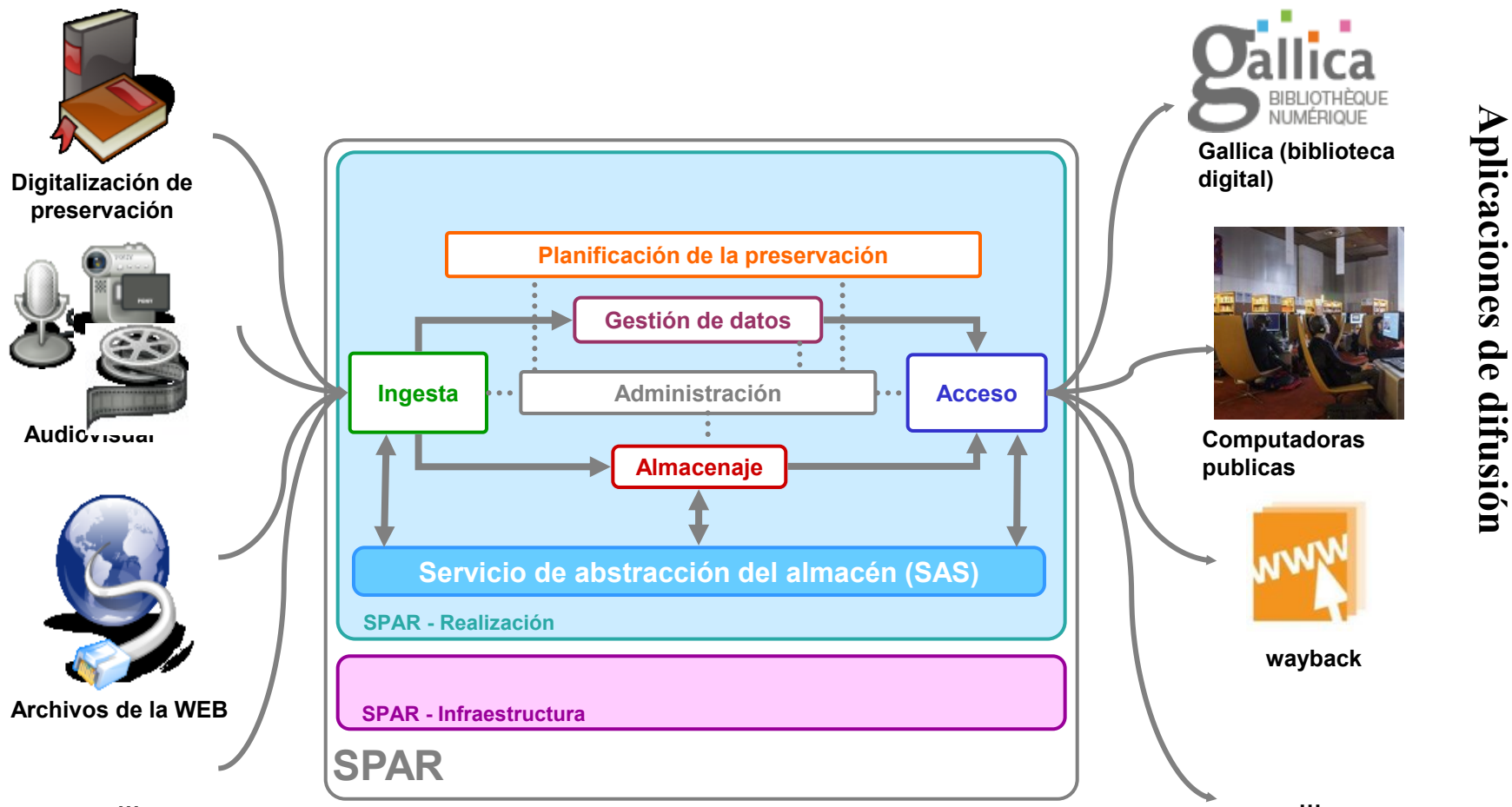
Deposito legal de la “web francesa”



Variedad de las colecciones.



Preservación digital en la BnF: SPAR



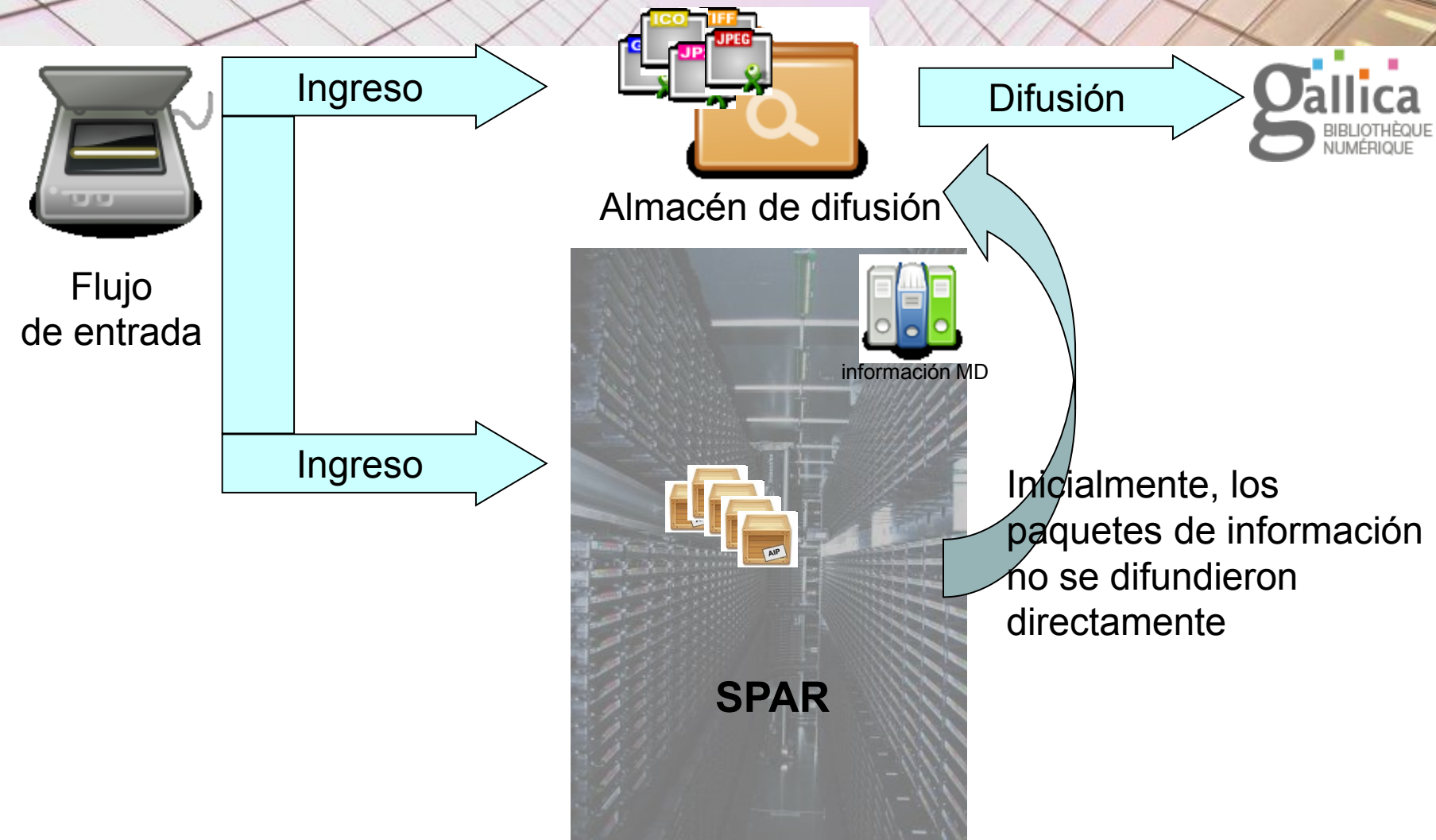
Unos datos sobre SPAR (el 22/01/2019)

- Inicio de la producción: mayo del 2010
 - Flujo corriente del programa de digitalización masiva de la BnF
 - Almacenamiento distribuido en 2 sitios.
- 9 055 476 paquetes de archivo
- 386 830 827 objetos de datos (ficheros)
- > 3.7PB (=3 785TB) de datos en bruto
- > 5.2 mil millones triples RDF (metadatos elementales)
- Ingreso de manuscritos muy voluminosos (paquetes con tamaño superior a los 100GB)
- Dos migraciones tecnológicas ya realizadas en 2012 y 2017: nueva generación de cintas
- Caminos disponibles: monografías, publicaciones periódicas, imágenes, audio, video, archivos de la web, documentos de oficina, libros electrónicos, archivos para terceros

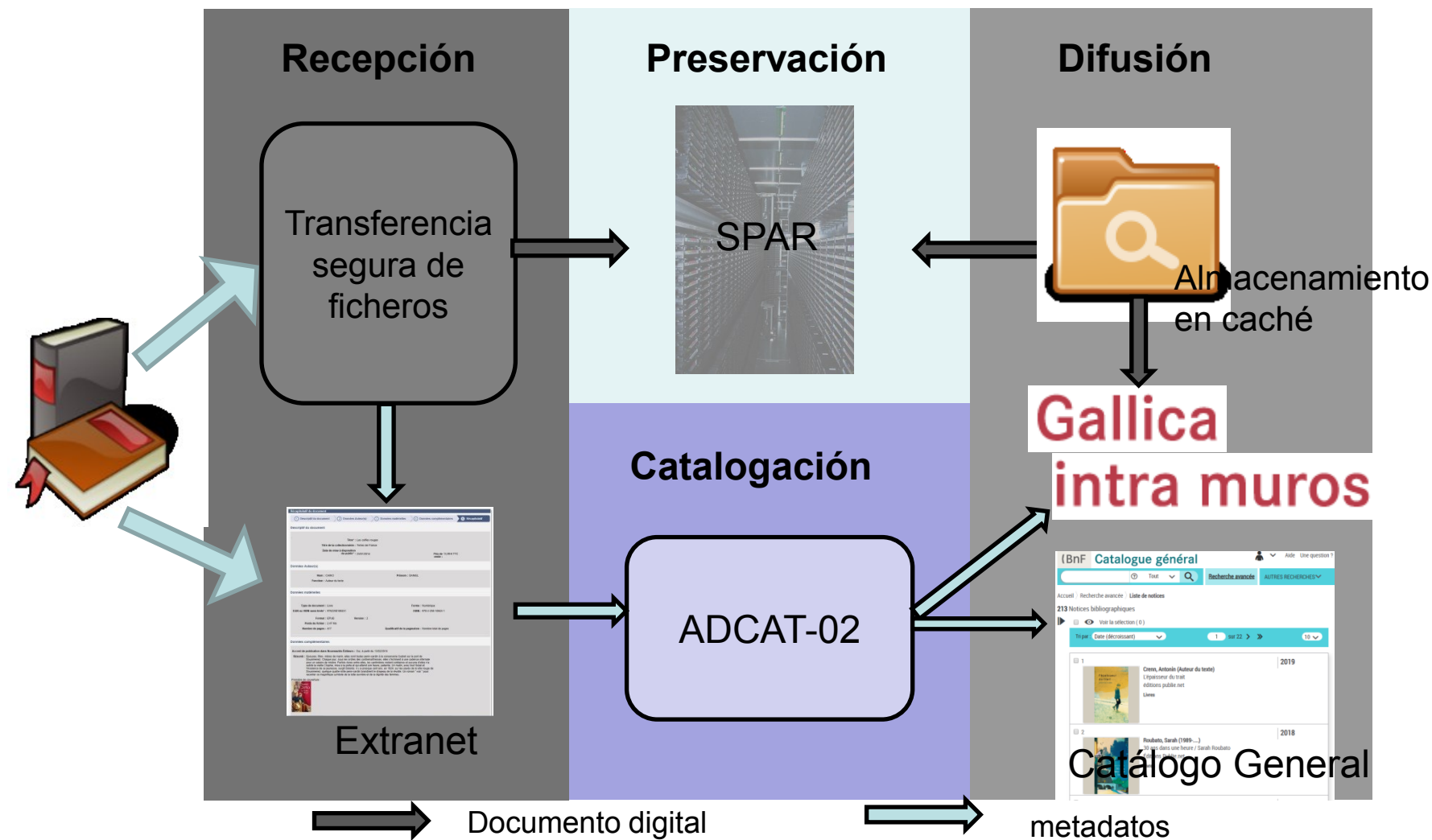
Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

Inicialmente, una caja fuerte digital



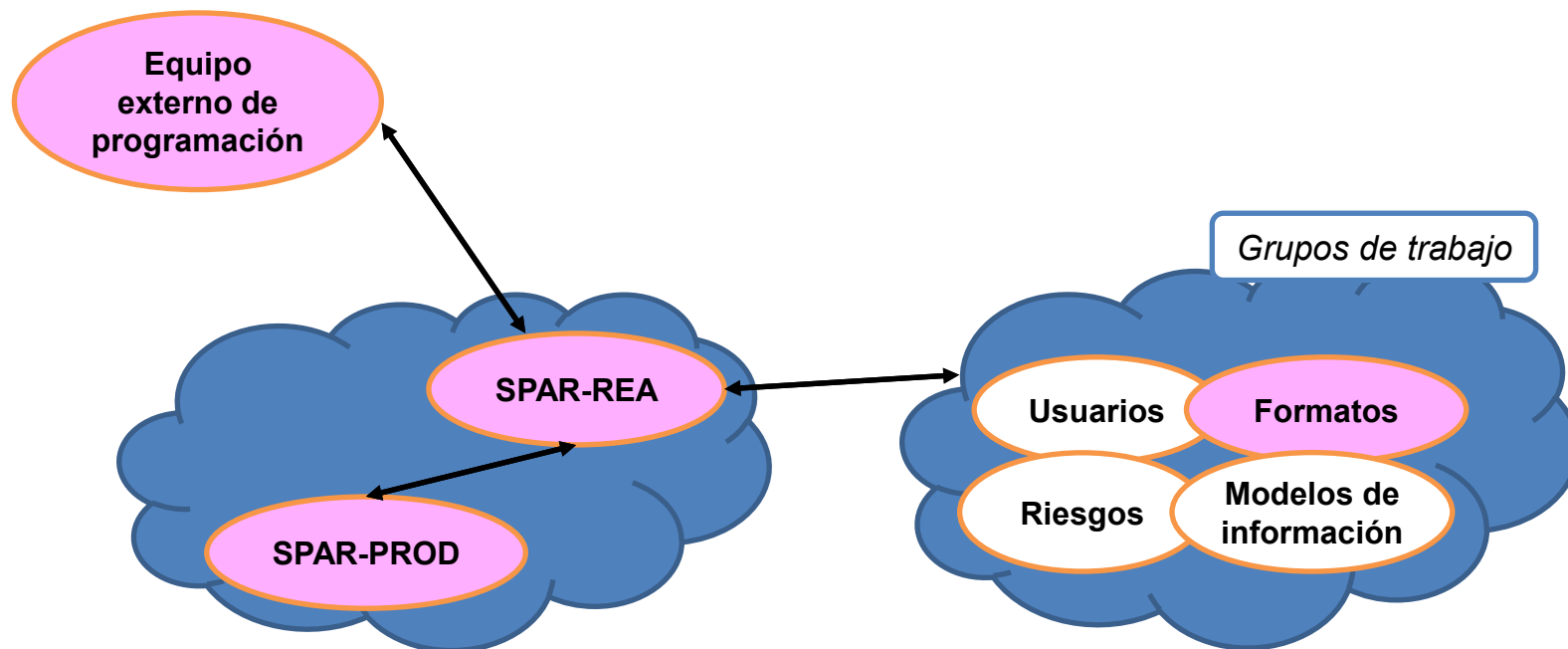
Ahora, depósito digital



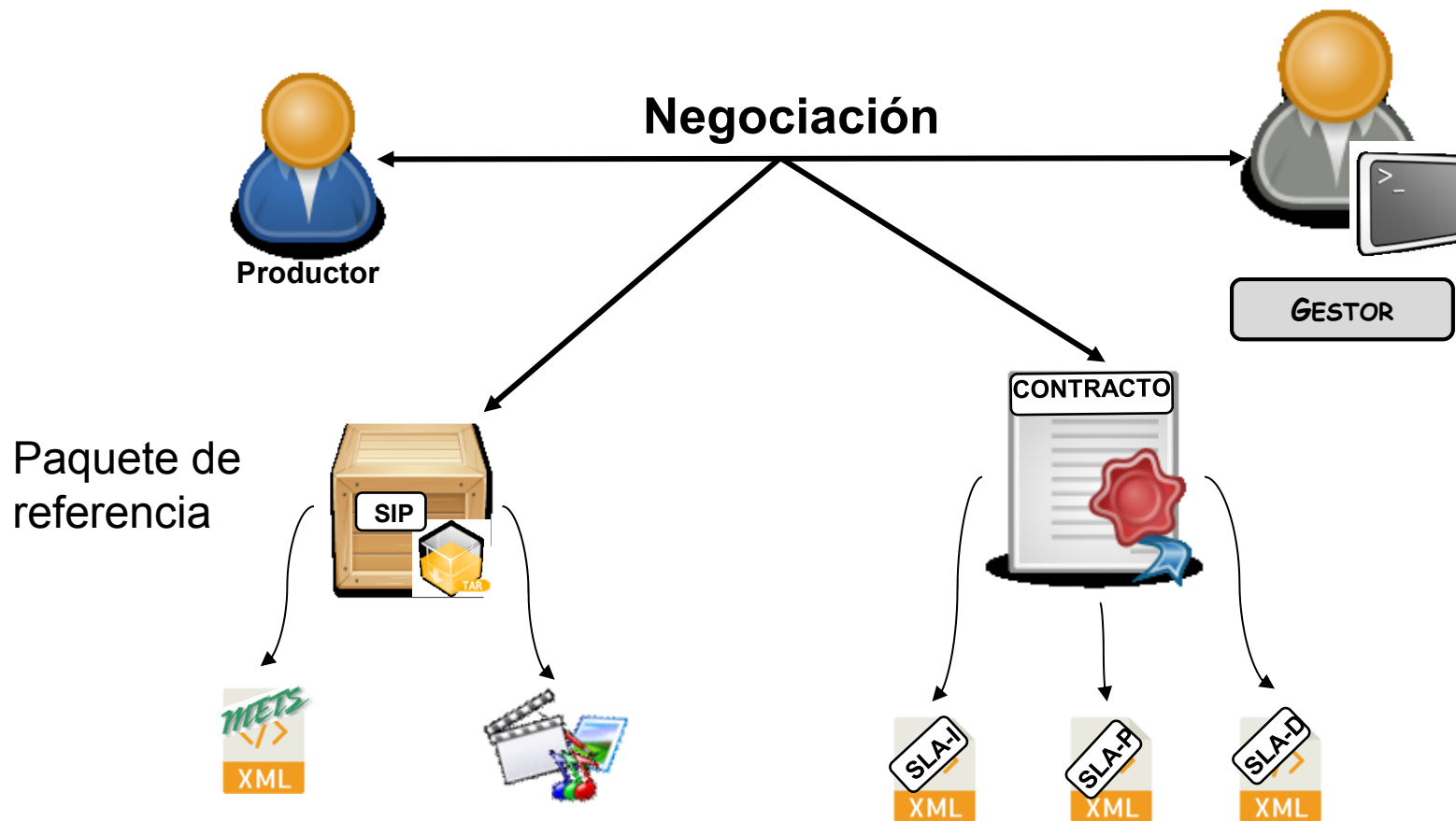
Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

Organización inicial del proyecto



Negociación por circuito: un contrato



Evolución en la organización

La biblioteca está construyendo una nueva organización para interactuar con el Archivo:

- censo de expertos en preservación digital (experticia en todas partes de la biblioteca)
- cada “circuito” tiene designado un gestor de la profesión, para negociar los SLA y definir las prioridades
- el departamento de informática es responsable del “deposito digital”

Lo más importante: la preservación ahora es parte integral de cada nuevo proyecto digital.

Personas involucradas en la planificación de la preservación



GESTOR DE
CIRCUITO



GESTOR



EXPERTO DE
PRESERVACIÓN



PROGRAMADOR

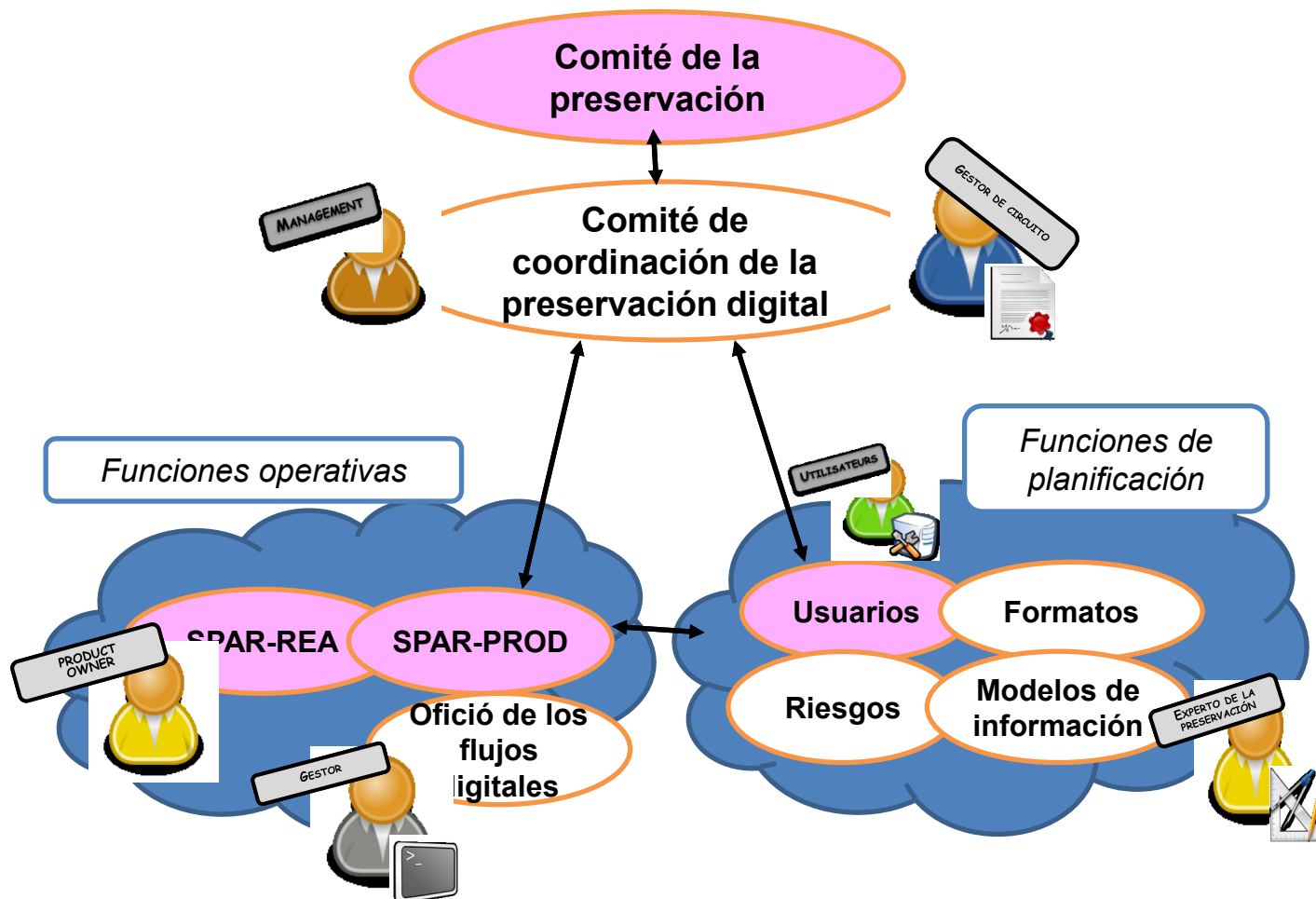


GESTOR
DOCUMENTAL

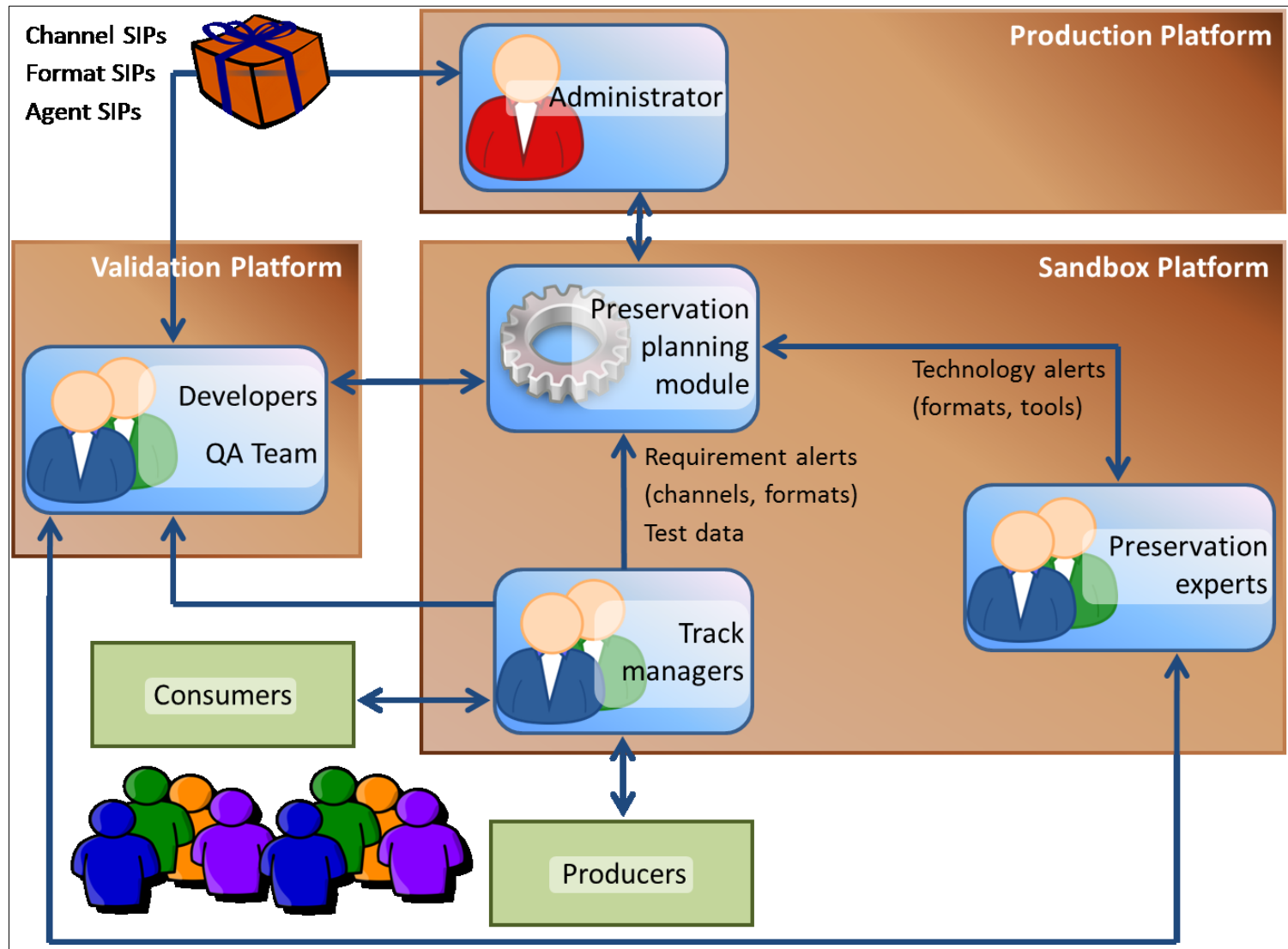


RIESGOS Y PLAN DE
EMERGENCIA

Organización de la preservación



La preservación como un proceso central



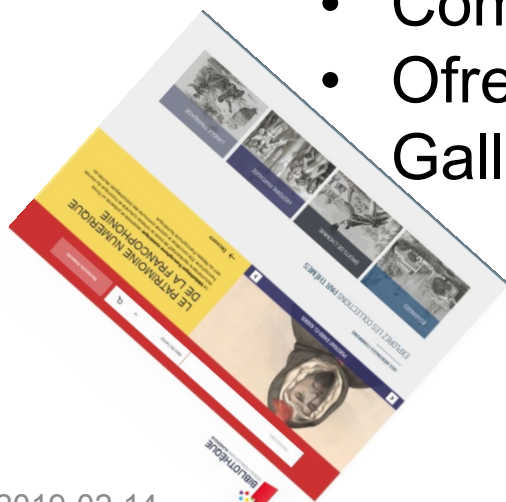
Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

Colaboración con otras bibliotecas



- Compartir la digitalización
- Integrar materiales ya digitalizados
- Preservar los materiales digitales en su nombre
- Compartir la difusión con Gallica
- Ofrecer un diseño propio: 'Marca blanca Gallica'



Oferta « BnF – Archivage Numérique »



- **Arreglos planeados para la integración, monitoreo y recuperación de los datos:**
 - ✓ Hecho a medida por los equipos de la BnF
 - ✓ O por el extranet « Espace Coopération » de la BnF
- **Medios proporcionados** dependiendo de la necesidad:
 - Sobre 2 cintas: acceso puede durar unas horas
 - Sobre 2 cintas y 1 disco: acceso en minutos, disponibilidad durante las horas de trabajo
 - Sobre 2 cintas y 2 discos: acceso en minutos, disponibilidad 7d/7.
- **Un contrato estándar** define los acuerdos de nivel de servicio (volumen, duración, formatos, metadatos, reglas, ...)

Costes del archivo de terceros

- Es una participación, al mejor costo, en el espíritu de compartir.
- Las listas de precios estándar se basan en:
 - La forma de descargar los datos a archivar (personalizados o extranet)
 - El tipo de media (cintas, discos).
 - El volumen a almacenar (de 1 a 50 TB).
 - La duración del contrato inicial (3, 5 o 8 años)
- El costo se deriva del nivel de servicio del acuerdo.



Escalabilidad en la adquisición de metadatos

(BnF)

- Entrega directa de los editores con flujos de trabajo B2B dedicados

Notice bibliographique

Notice Au format public ▼



Type(s) de contenu et mode(s) de consultation : Texte : électronique

Auteur(s) : Pautrel, Marc

Titre(s) : La vie des écrivains classiques [Texte électronique] : le roman, la lecture, la biographie et l'écriture : réflexion côté table de travail et de l'écrivain dans le rôle de celui qui fait entrer les paquebots en bouteille / Marc Pautrel

Publication : [Montpellier] : publie.net, 2009

Description matérielle : 1 ressource dématérialisée

Détail : 1 fichier : EPUB

Collection : Ouvrez...

Note(s) : Métadonnées fournies par l'éditeur

Numéros : ISBN 978-2-8145-0037-2 : 0,99 EUR

EAN 3612220342081 (fichier EPUB)

EAN 9782814500372 (multiformat)

Notice n° : FRBNF45640197 (notice reprise d'un réservoir extérieur)

Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

Aspecto económico: volumen



Digitalización inicial
NB - 300dpi - TIFF G4
1 pagina ~ **200KB**

Digitalización corriente
Color (24 bits) – 400dpi
– TIFF sin compresión

Digitalización corriente
Color (24 bits) – 400dpi
– **JPEG 2000**

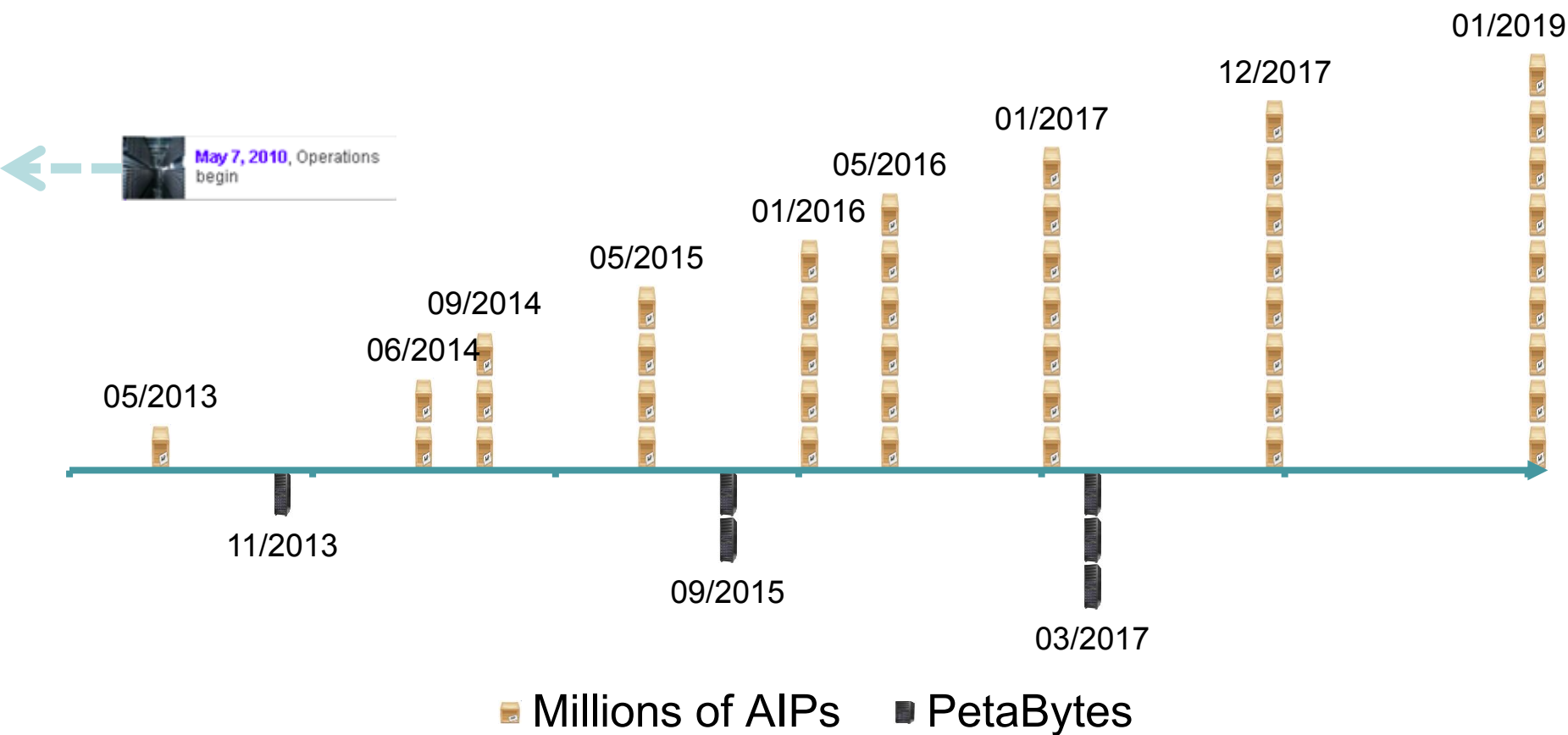
Multiplicado por 400 !!!

Dividido por 4

1 pagina ~ **80MB**

1 pagina ~ **20MB**

Cronología del crecimiento



Aspecto económico: infraestructura de almacenamiento

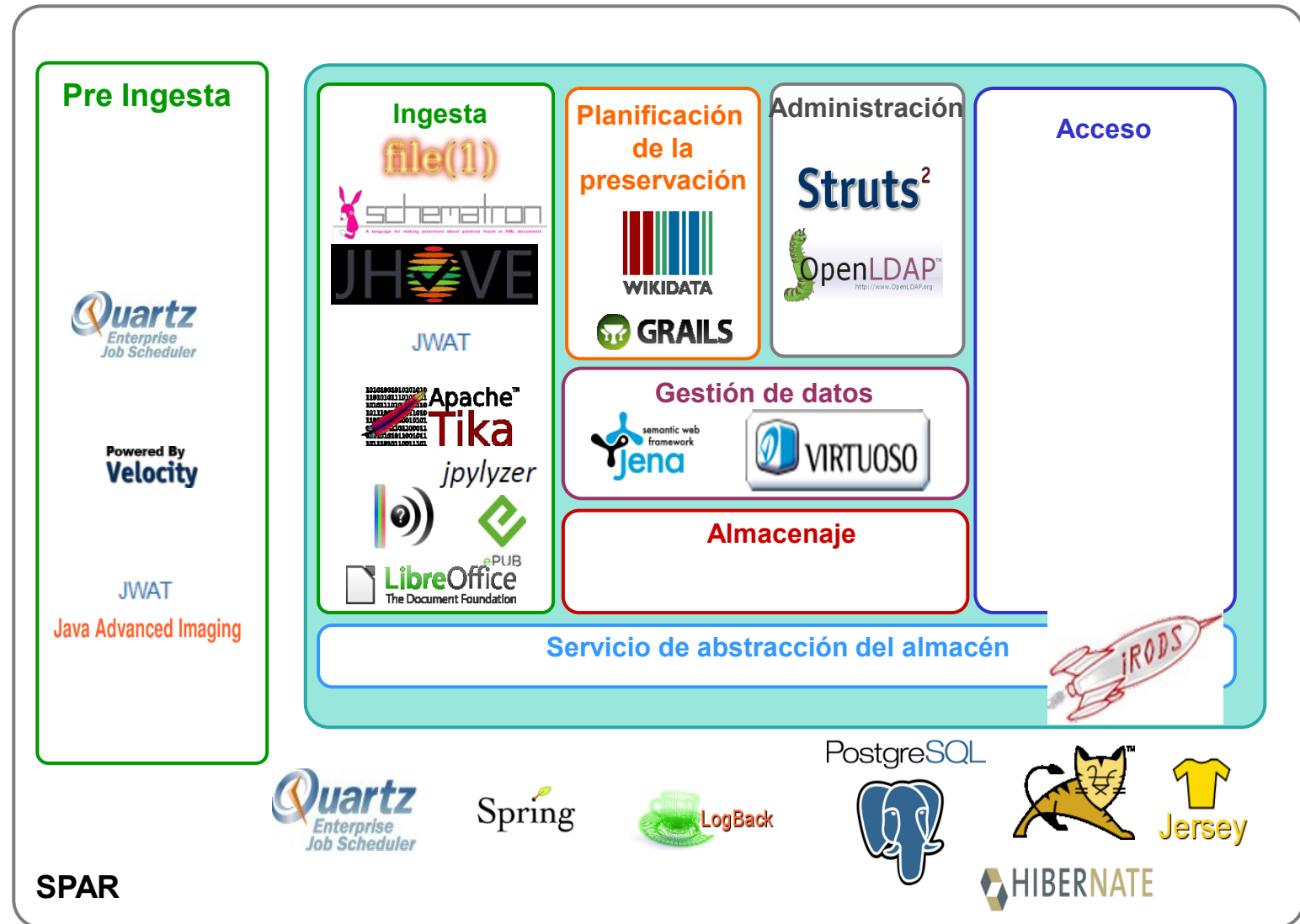
El “Cloud” aún no es una opción: problemas legales, costo recurrente, problema de transferencia de datos cuando se pasa de un proveedor a otro (plan de escape)

Disco duro solo para la difusión.

Las cintas siguen siendo económicamente viables (10 veces menos costosas que el disco duro) y eficientes energéticamente



Aprovechar el esfuerzo de la comunidad {BnF de código abierto



Orden del día

- Contexto de los activos digitales en la BnF
- Desde archivar hasta preservar
- Establecer una nueva organización
- Compartir datos y flujos de trabajo
- Arreglárselas con el volumen
- Conclusión

Conclusiones

Para que la preservación digital sea una actividad sostenible, necesita:

- Implicación de la institución (misión central, conocimientos, formación,...)
- Compartir con todo el ecosistema
- Construir un plan de negocios
- Y siempre, anticipar los riesgos



Gracias por su atención

¿Preguntas?

thomas.ledoux_at_bnf.fr

{BnF