

## Seminario EDCine/Archives

Tecnologías Digitales en los Archivos Fílmicos  
Ingesta, almacenamiento, acceso y distribución digitales.

# Acceso a colecciones digitalizadas

Valencia, 4, 5 y 6 de marzo, 2009

1 bit es 0 o 1

8 bits = 1 Byte = 256 valores distintos

## ESTOS SE UTILIZAN PARA EL TAMAÑO/ESPACIO EN DISCO

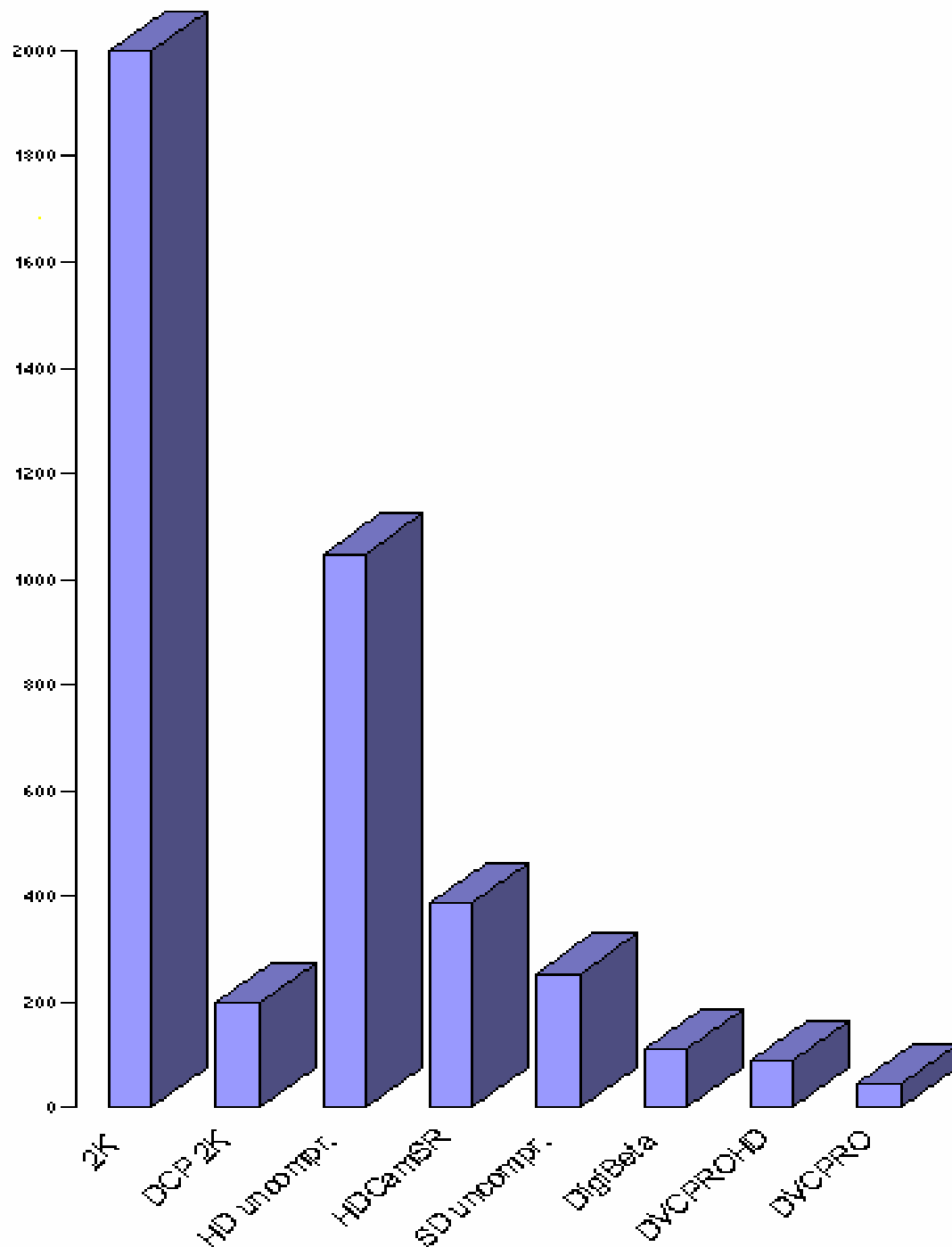
1.024 bytes	= 1 Kilobyte (KB)	= $2^{10}$ bytes
1.024 Kilobytes	= 1 Megabyte (MB)	= $2^{20}$ bytes
1.024 Megabytes	= 1 Gigabyte (GB)	= $2^{30}$ bytes
1.024 Gigabytes	= 1 Terabyte (TB)	= $2^{40}$ bytes
1.024 Terabytes	= 1 Petabyte (PB)	= $2^{50}$ bytes

## PERO PARA EL BIT RATE (MEDIR EL TRANSPORTE) USAMOS

‘bits por segundo’ o *bps* o *bit/s*, y sucesivamente: Kilobits, Megabits etc.)’

Así tenemos: **Kbps** (Kbit/s), **Mbps** (Mbit/s), **Gbps** (Gbit/s)

Ejemplo: “ *DSL has a bit rate of 256kbps* “



## Formatos: ¿qué tamaño?

**Dimensiones en GB  
de ficheros para una  
película de 2 horas**

<b>2K</b>	<b>2000</b>
<b>DCP 2K</b>	<b>~200</b>
<b>HD sin compr.</b>	<b>1048</b>
<b>HDCamSR</b>	<b>388</b>
<b>SD sin compr.</b>	<b>254</b>
<b>Betacam Digital</b>	<b>110</b>
<b>DVCPROHD</b>	<b>88</b>
<b>DVCPRO</b>	<b>44</b>

# Formatos: ¿qué tamaño?

		Transport	Storage
	<i>Formats</i>	<i>bit rate in Mbps</i>	<i>GB/hr</i>
	2K	2.300	1.010
	DCP	<250	~100
HD	Uncompr	1.200	524
HD	HDCamSR	440	194
HD	HDD5	320	140
SD	Uncompr	288	127
SD	DigiBeta	125	55
HD	DVCPRO100	100	44
SD	DVCPRO	50	22
SD	DVCam	25	11

Parámetros preferentes / principales:

- Calidad / finalidad
- Transmisión
- Almacenamiento (incluido gastos)

HDCam (SR), un estándar “de-facto”

HD DV bajo presupuesto, noticias

Betacam Digital un estándar “de-facto” para SD

DV noticias, prosumer

2k / 4k –DI sólo

DCP

## Ethernet / Redes internas

10Mbps / 100 Mbps / 1Gigabitps / 10Gigabitps

## (Banda ancha) Internet

dial-up	28,8 / 56,6 kbps
cable	256 / 512 kbps
ISDN	max 2 Mbps
ADSL	1,5 - 8 Mbps
ADSL2	5 - 28 Mbps
VDSL	Max 52 Mbps
VDSL2	50 - 100 Mbps

DVB-T = 5 – 32 Mbps

DVB-S for SD = 4 – 8 Mbps

DVB-S for HD = 15 – 20 Mbps

		Transport	Storage
	<i>Formats</i>	<i>bit rate in Mbps</i>	<i>GB/hr</i>
	2K	2.300	1.010
	DCP	<250	~100
HD	Uncompr	1.200	524
HD	HDCamSR	440	194
HD	HDD5	320	140
SD	Uncompr	288	127
SD	DigiBeta	125	55
HD	DVCPRO100	100	44
SD	DVCPRO	50	22
SD	DVCam	25	11

**Ethernet / Internal networks:**  
 10Mbps / 100 Mbps / 1Gigabit ps /  
 10Gigabit ps

ISDN	max 2 Mbps
ADSL	1,5 - 8 Mbps
ADSL2	5 - 28 Mbps
VDSL	Max 52 Mbps
VDSL2	50 - 100 Mbps
DVB-T	5-32 Mbps
DVB-S SD	4-8 Mbps
DVB-s HD	15-20 Mbps

**¿Entonces?**

**COMPRESOR, COMPRESOR**

## Compresión:

- Lossless - Sin pérdida – compresión matemáticamente sin pérdida - máx ~2,5 veces
- lossy compression – compresión con pérdida – nivel de compresión casi ilimitado
- ‘visually lossless’ - ‘sin pérdida visual = con pérdida (e.j: formatos de vídeo digital)

## Compresión:

- Imágenes que se pueden modificar después de la compresión (montaje, restauración,...)
- Imágenes que **NO** se pueden modificar tras la compresión

Estrategias para reducir la dimensión (y por consiguiente la velocidad de transmisión en streaming)

- Comprimir los datos en la imagen (ej.: redundancia)
- Reducir la resolución
- Reducir el bit-depth
- Recrear / Interpolar las imágenes (ej: archivar sólo diferencias entre imágenes)

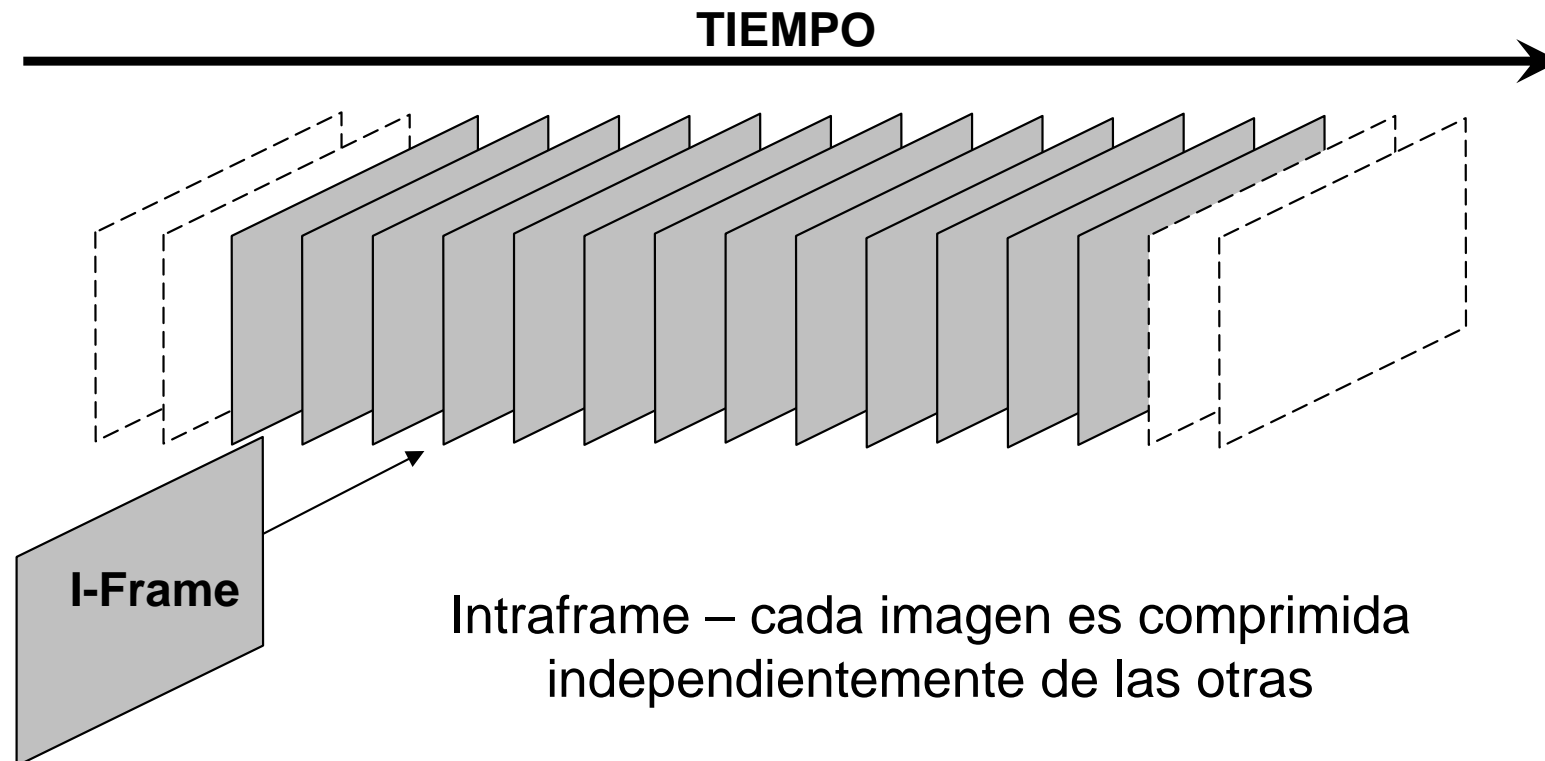


Estrategias para reducir la dimensión (y por consiguiente la velocidad de transmisión en streaming)

- Comprimir los datos en la imagen (ej.: redundancia)
- **Podría impedir el procesado**
- Reducir la resolución
- **Podría impedir su uso / que sea apreciado**
- Reducir el bit-depth
- **Podría impedir su manipulación / restauración**
- Recrear / interpolar las imágenes
- **Podría impedir la edición y manipulación**

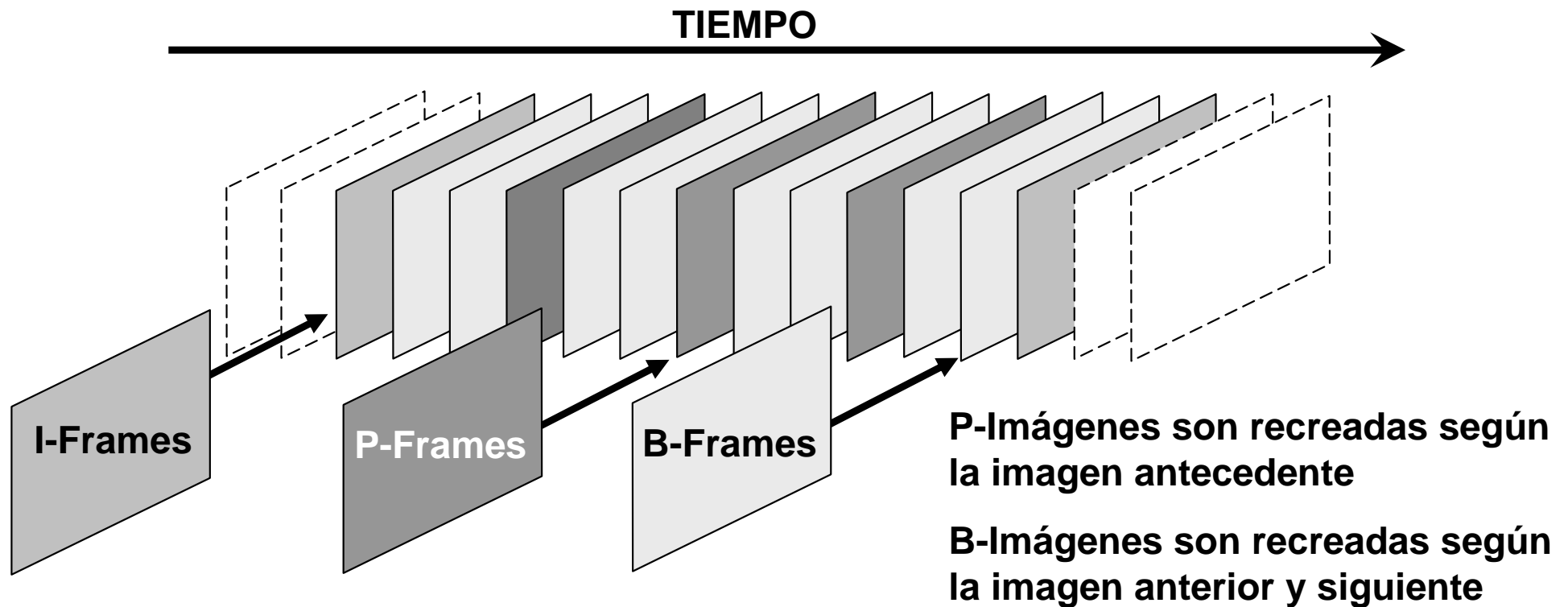
Compresión:

Imágenes que se pueden manipular tras la compresión

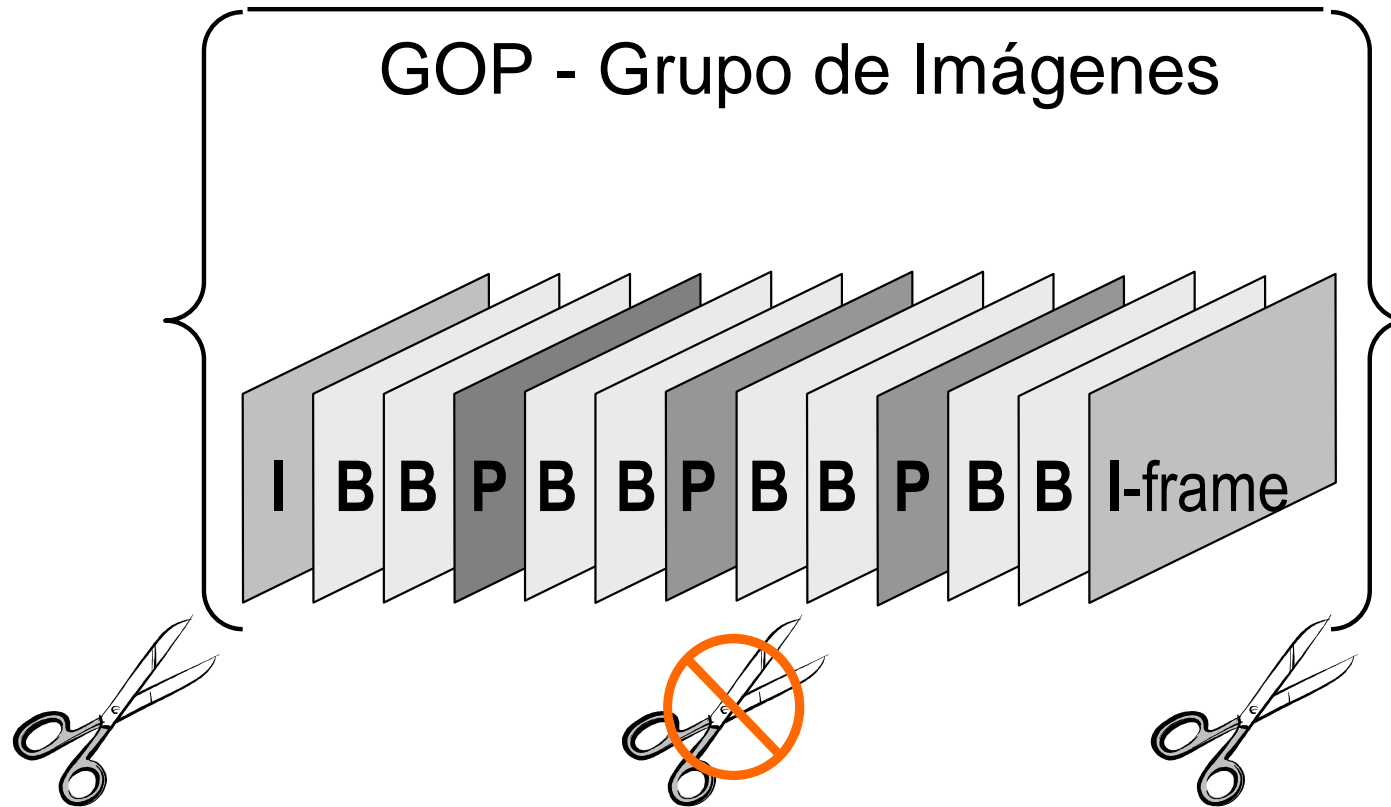


Compresión:

Imágenes que NO se pueden manipular tras la compresión



# Compresión



## Estrategias para reducir la dimensión

- Comprimir los datos en la imagen (ej. redundancia)
- Reducir la resolución
- Reducir el bit-depth
- Recrear / Interpolar las imágenes (ej. archivar sólo diferencias entre imágenes)

**...todavía, podría NO ser suficiente**

SD no-comprimido 270Mbps / cable 256 Kbps  
**= la proporción es 1054:1**

**necesitamos más:**

- **Cambiar la dimensión de las imágenes: de 4k a 2k a 1920x1080 a 720x480 a 320x240 a 192x144 a 160x120**
- **Reducir el framerate: de 25 i/s a 15 i/s a 5 i/s**

# Compresión

<i>2 minutos</i>	<i>3000 imágenes</i>		<i>dimension en KB</i>	<i>Mbps</i>
augenblick test max.avi	<a href="#">link</a>		421.947,0	<b>27,47</b>
DVD Format - VOBfile	<a href="#">link</a>		116.356,0	<b>7,58</b>
augenblick H264 25f 50%.mov	<a href="#">link</a>		112.931,0	<b>7,35</b>
augenblick 320x240 3,5Mbps.wmv	<a href="#">link</a>		53.170,0	<b>3,46</b>
augenblick 320x240 900Kbps.wmv	<a href="#">link</a>		13.881,0	<b>0,90</b>
augenblick 320x240 500Kbps.wmv	<a href="#">link</a>		7.243,0	<b>0,47</b>
augenblick 320x240 for 255max.wmv	<a href="#">link</a>		3.935,0	<b>0,26</b>
augenblick 320x240 for 128max.wmv	<a href="#">link</a>		1.373,0	<b>0,09</b>

	<i>Bitrate</i>		<i>Bitrate</i>
dial-up	<b>28,8 / 56,6 kbps</b>	ADSL2	<b>5 - 28 Mbps</b>
cable	<b>256 / 512 kbps</b>	DVB-T	<b>5-32 Mbps</b>
ISDN	<b>max 2 Mbps</b>	DVB-S SD	<b>4-8 Mbps</b>
ADSL	<b>1,5 - 8 Mbps</b>	DVB-s HD	<b>15-20 Mbps</b>

¿Qué son estos ficheros?

Codificación propietaria, basada sobre estándares pero 'modificados'

- Windows Media
- QuickTime
- RealNetworks

Son variantes, más o menos directas de

MPEG-1 – el primero, para multimedia y CD

MPEG-2 – para broadcast, DVB, DVD

MPEG-4 – el jovencito de la familia

H.264 (MPEG-4 AVC) – de la TV al móvil, al Blu-ray.

Tenemos tres opciones principales:

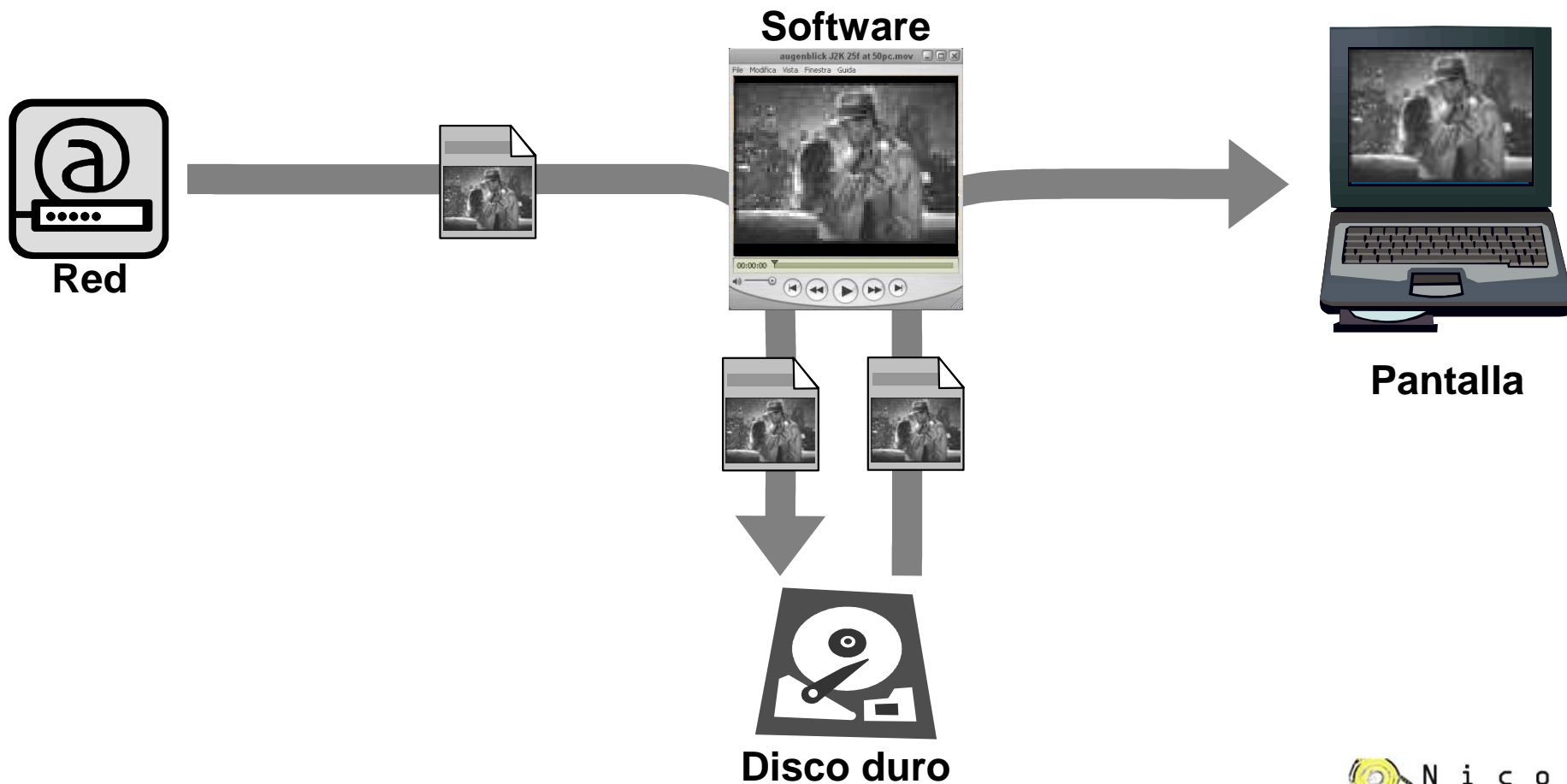
Internamente: red, elementos físicos

Externamente:

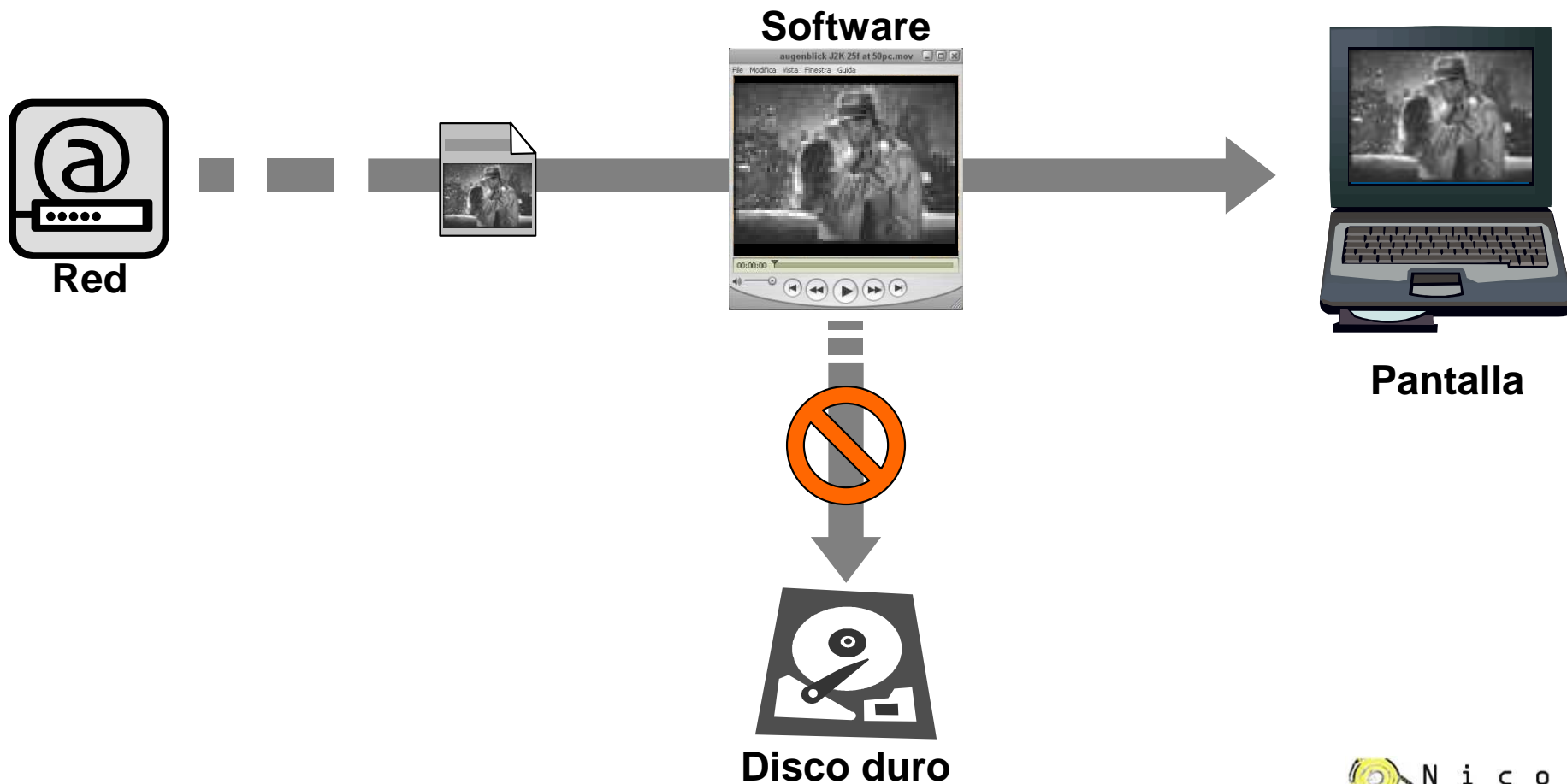
- Entrega física – Enviar la cinta (formato de vídeo) o el fichero (en algún soporte: DVD, cinta de datos, disco duro, etc.)
- -Descarga de ficheros/reproducir después de descargar
- -Streaming



## 2. Descarga (& reproducción)



## 3. Streaming



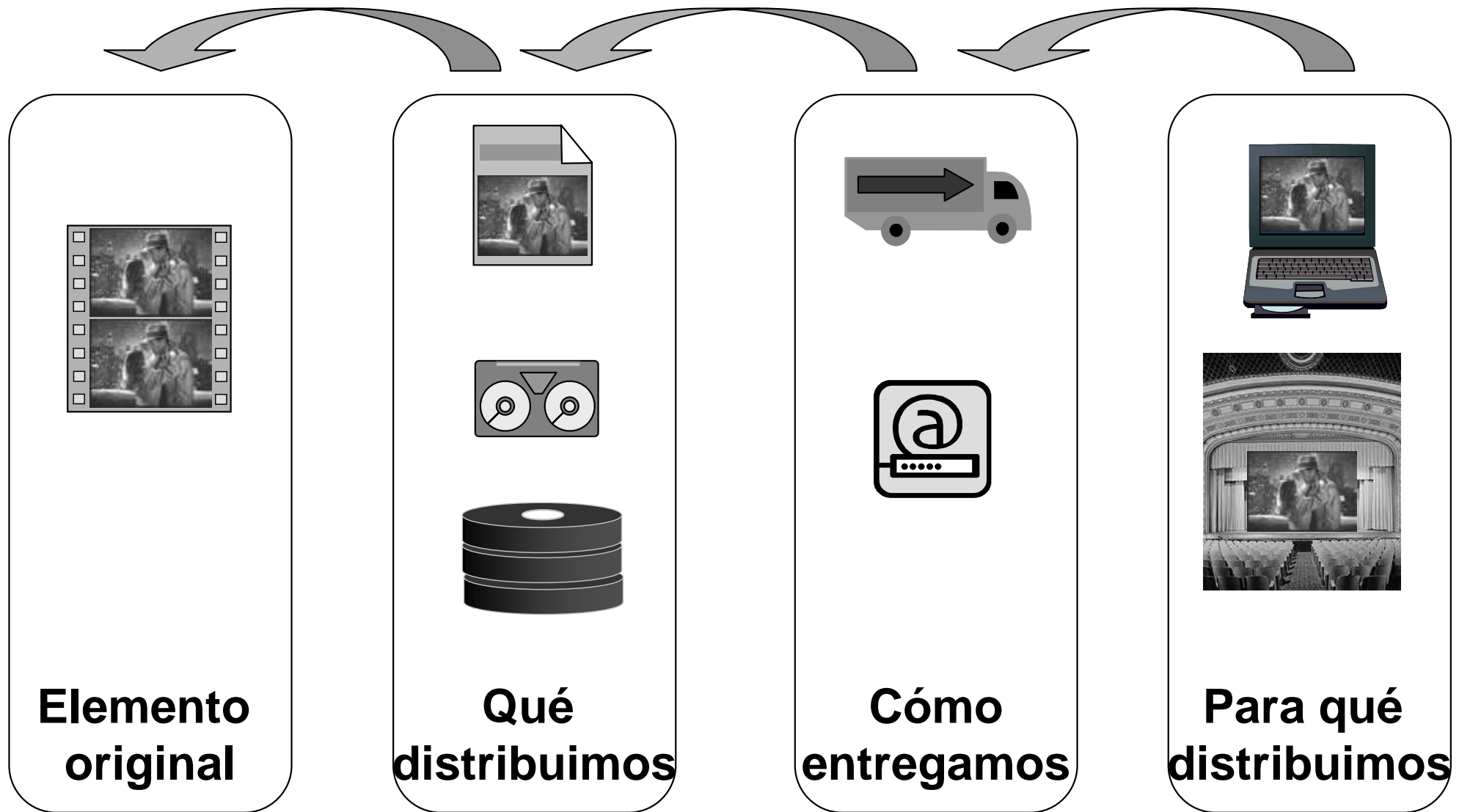
Todos los formatos propietarios **están compitiendo por streaming**

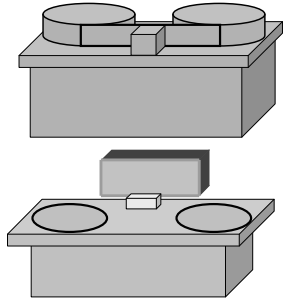
## Una difícil elección

- Predecir el rendimiento de la red (variable, QoS-calidad de servicio no garantizada)
- Predecir la conexión del usuario (opción: VBR, el software adapta el bitrate al rendimiento de la red)
- Predecir la finalidad del uso (examinar o visionar, en directo o después de la descarga, etc.)
- Evaluar la penetración de cada ‘media player y software’ (WM, QuickTime, RealNetworks)
- Eligir un estándar para servidores web y streaming (están en competencia y no son compatibles: windows es sólo windows, Apple es sólo Apple, RealNetworks y Flash sirven a ambos, etc.)
- ‘Download and play’ no es crítico, puede pasarse NO en tiempo real, **– pero no se puede explorar!**

## Efectos colaterales y problemas:

- Problemas de almacenamiento
  - La dimensión de los ficheros para streaming no es un problema, pero:
  - La dimensión de los ficheros para la descarga (ej.: calidad MPEG2 DVD o Blu-ray, o Cine Digital) puede ser un problema
  - El número de versiones diferentes puede ser importante: bit rates diferentes, formatos diferentes, media players diferentes
- Problemas de proceso
  - Cuándo codificar – después de la codificación es difícil, si no imposible, modificar el contenido
  - Cuántos ficheros cada ‘título’?
  - Tiene que ser automatizado
  - Las exigencias cambian constantemente (la red siempre mejora, nuevos formatos, nuevas versiones del software, etc.)
- Problemas de estrategia
  - Tenemos que conocer/suponer quién quiere las imágenes, cómo, porqué, y tenemos que controlar los cambios
  - ...y todavía no hemos hablado del contexto!





## Calidad necesaria:

- Proyección?
- Navegación?
- Master TV?

*No se puede superar la calidad inicial*

**Cómo digitalizamos**

## Si SI:

Archivos/formatos a procesar, baja o nula compresión, mayor bit rate,

## Si NO:

- Más tarde?  
Pueden hacerse 2 versiones o m  
- Nunca?  
Codificar directamente

**Procesar ahora?**

## Si NO:

Codificar directamente, opción para guardar en alta calidad para futuras versiones.

## Si SI:

Guardar en alta calidad, sin procesar y confiar que esté suficientemente bien

**Procesar después?**

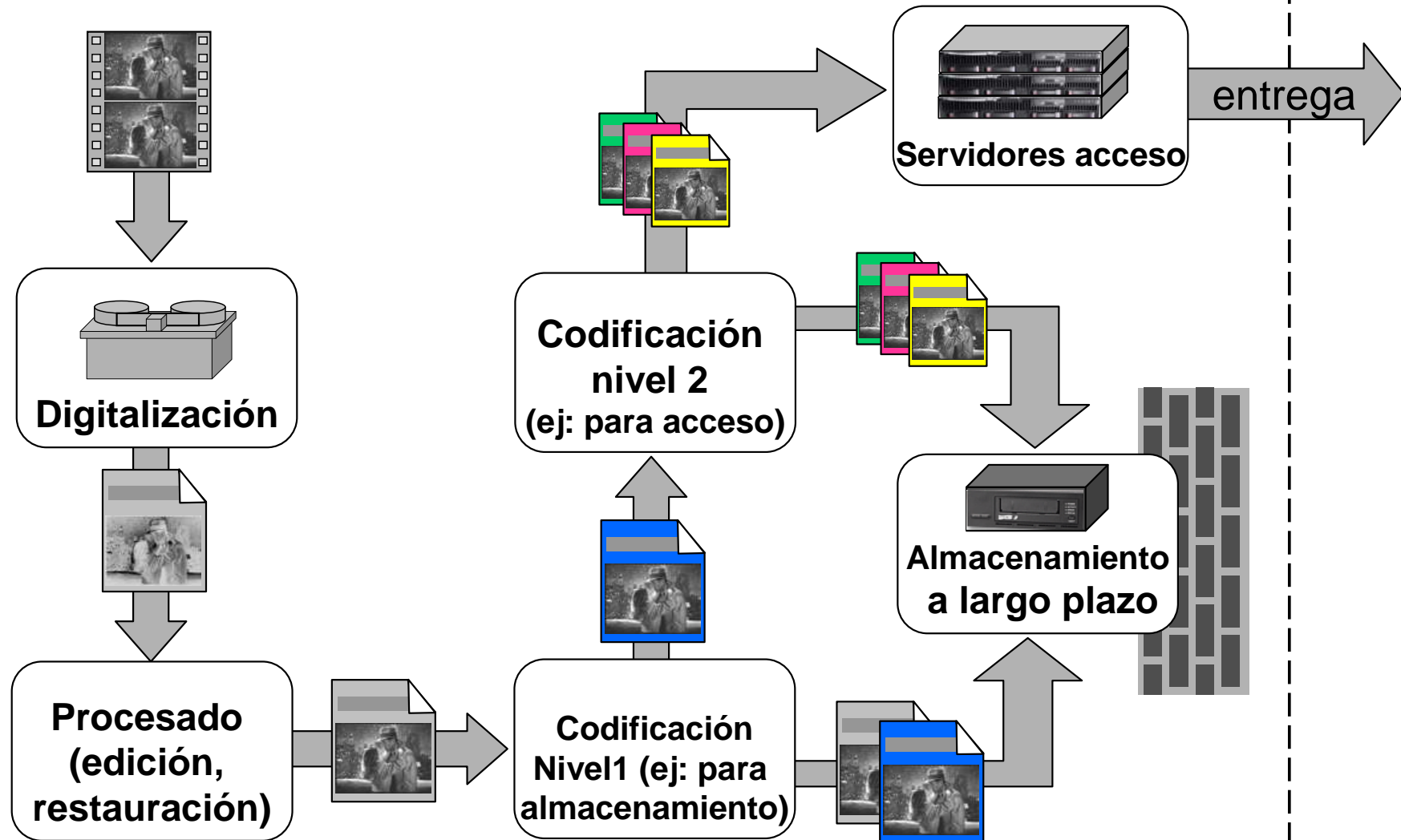
## Si sólo una entrega:

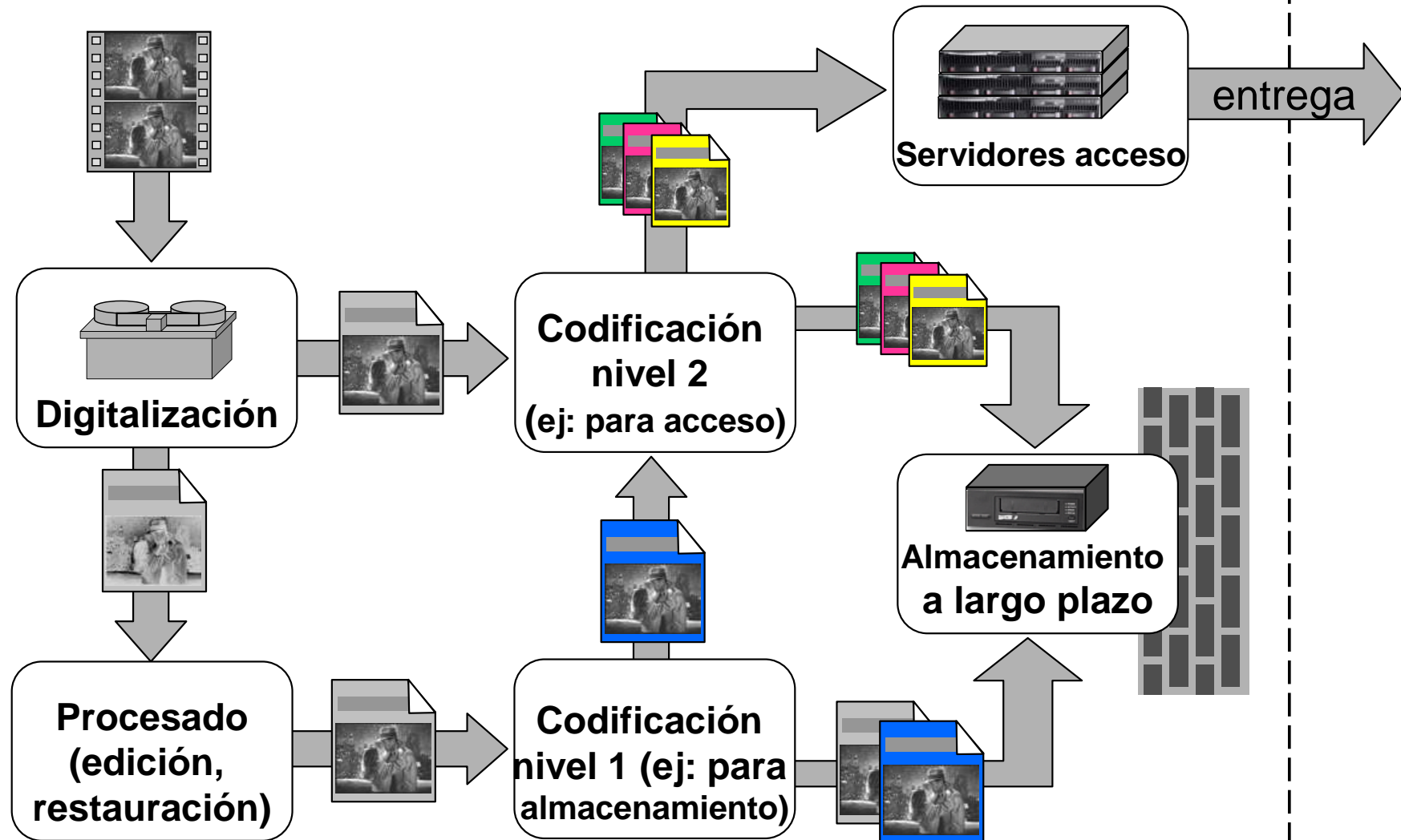
Codificar según se requiera y guardar para el futuro.

## Si son posibles más entregas:

- Codificar todo lo posible  
- Codificar sólo lo probable

**Cómo codificamos**





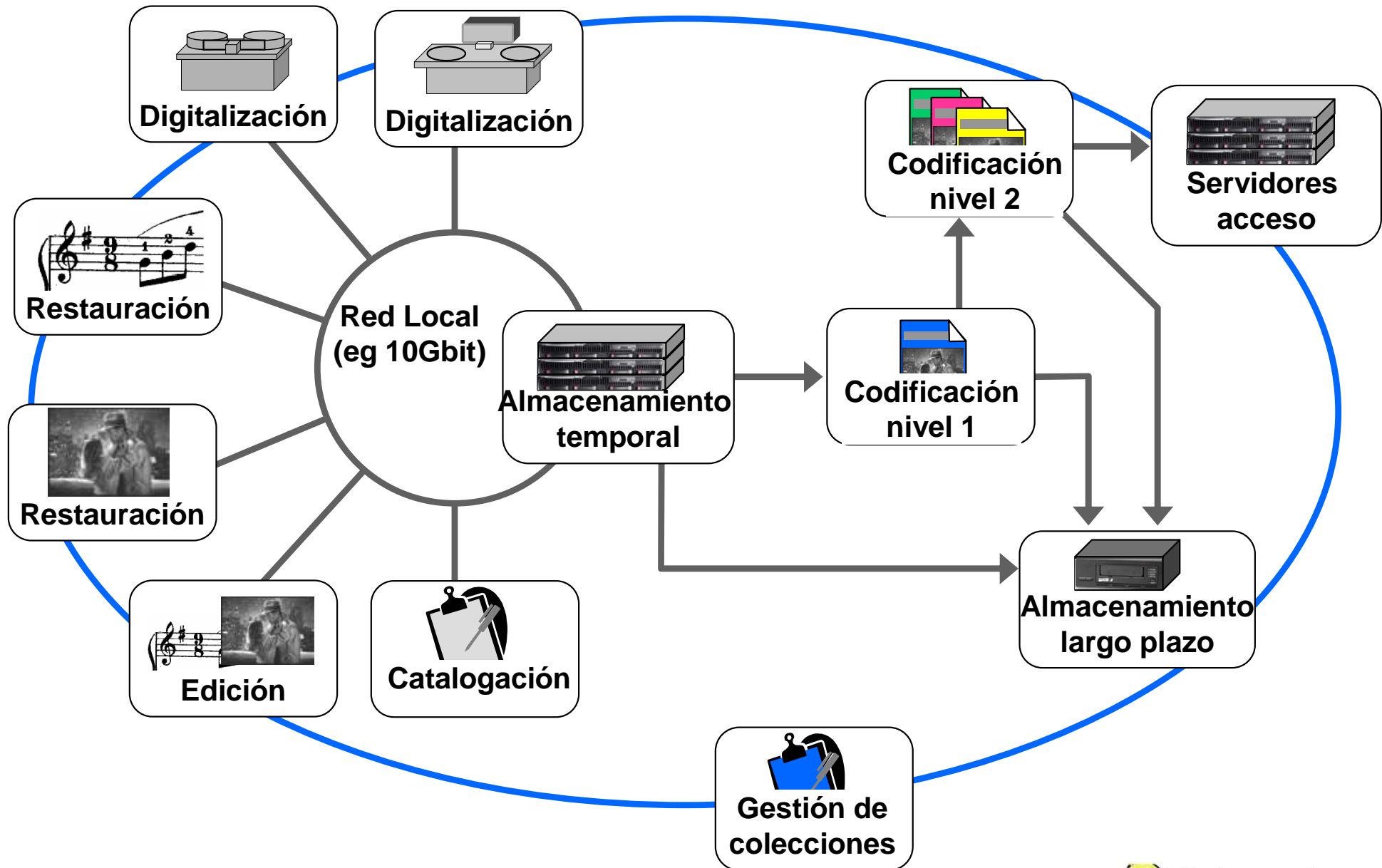


El diagrama ilustra un proceso que ejecuta un sólo paso a la vez, uno tras otro. Además, hay una estación de procesamiento por cada tipo (una para digitalización, una para procesamiento y edición, una para restauración, etc.)

Pero no es realista. Podría ser así al principio, pero no a lo largo del proceso.

La realidad es que hay múltiples procesos y algunos de ellos se desarrollan en paralelo.

Por tanto debemos mirar hacia un proceso de trabajo e infraestructuras diferentes...

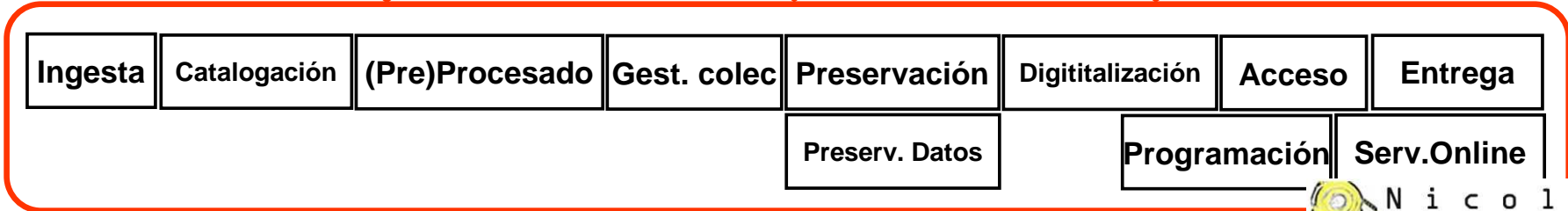


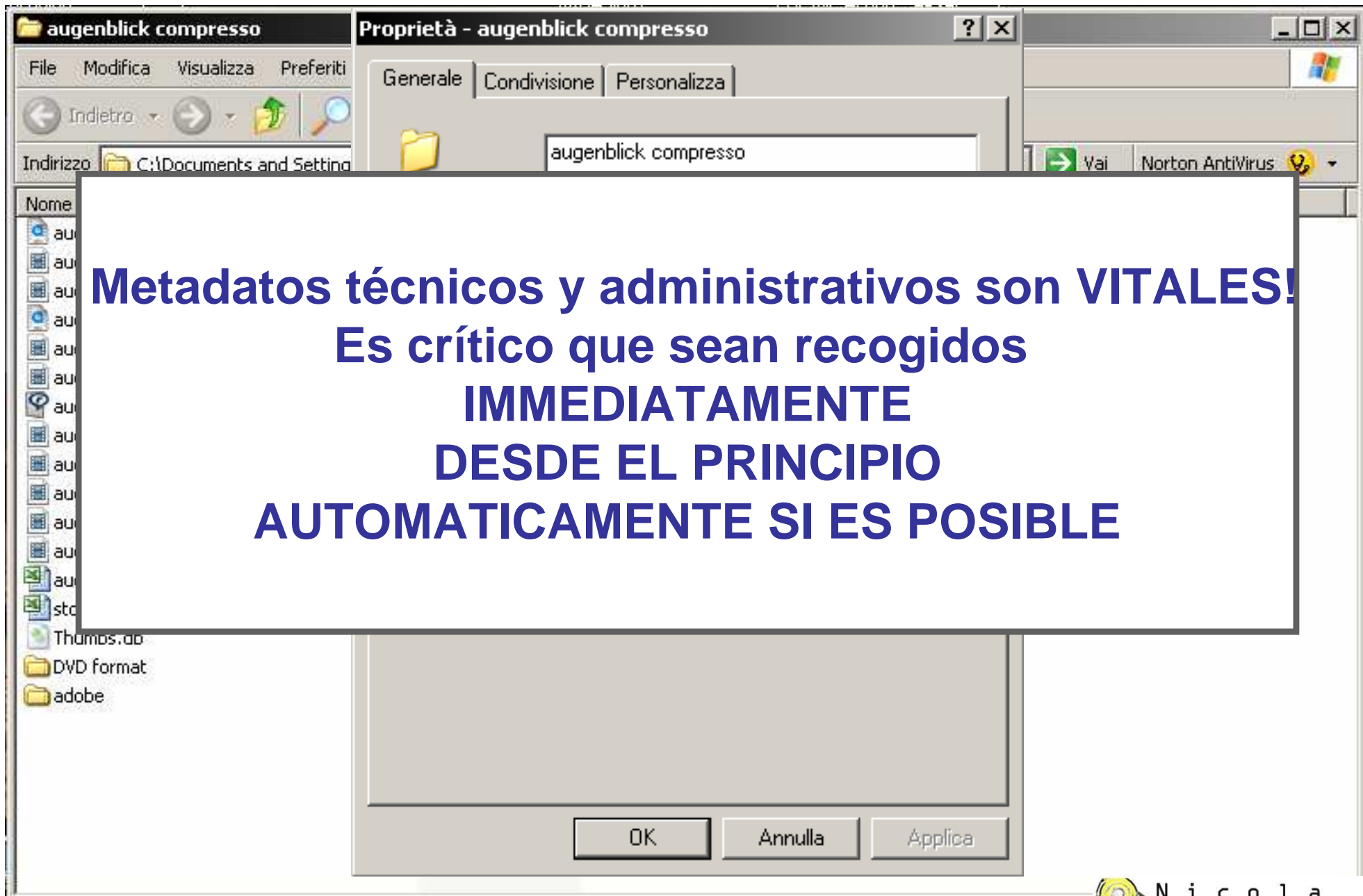
## Situaciones de la vida real:

- “Hey ¿quién decidió escanear el duplicado positivo? No está completo, sólo el duplicado negativo está completo.”
- “¿Por qué hiciste sólo el DVD y no el fichero Quicktime?”
- “Hey ¿quién acortó el plano nº 55? Ahora la banda de sonido restaurada ya no casa (acopla) con la película!”
- “Hey ¿quién codificó esta versión? ¡Todavía no me había hecho con la corrección de color.!”
- “Hey ¿dónde está la versión para streaming?”
- “¿Tú sabes qué fichero debería enviar para almacenar?”
- “¿Alguien sabe dónde ha ido a parar el fichero con que estaba trabajando yo ayer?”

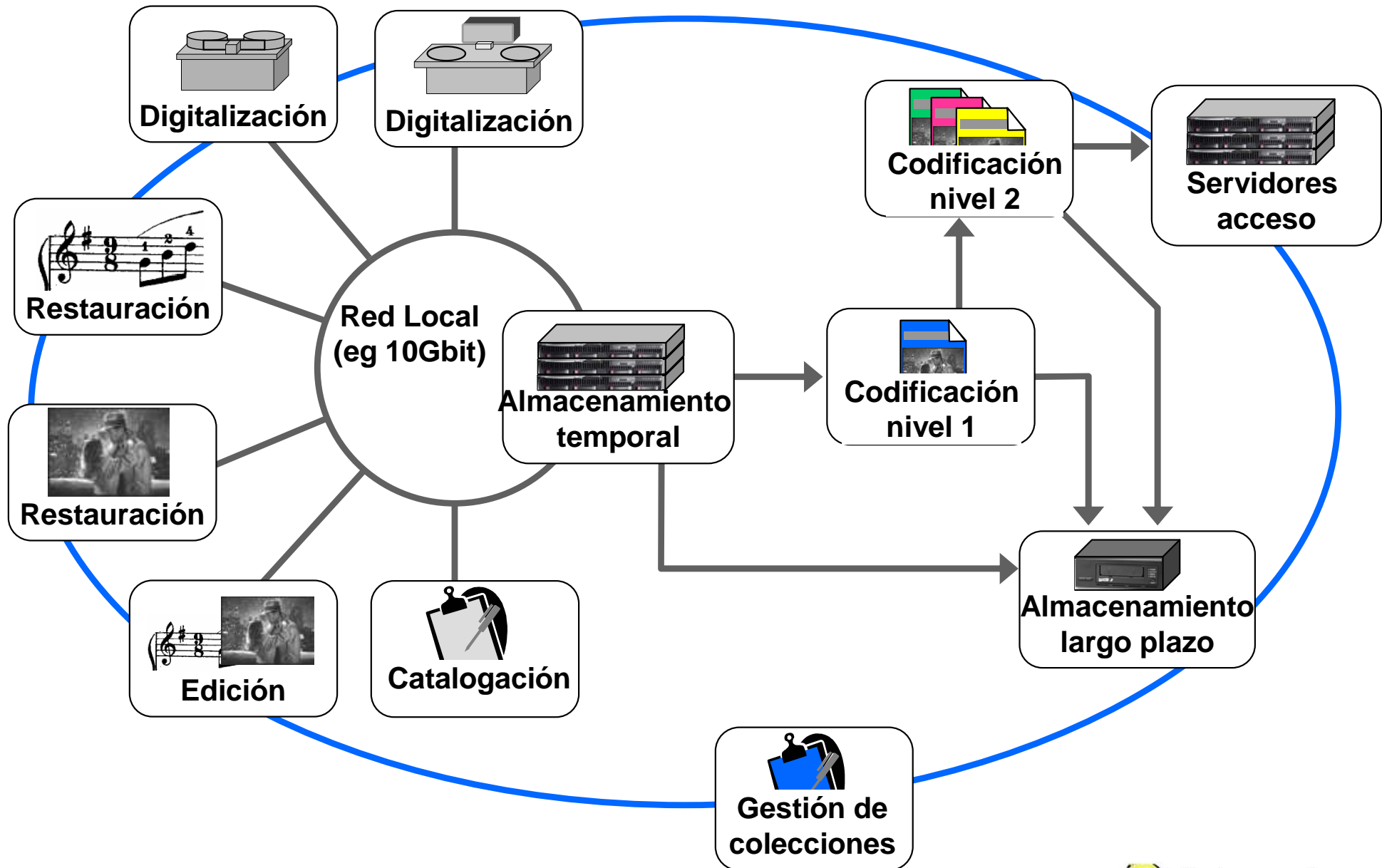
Colecciones	Procesado	Metadatos	Acceso	
Fílmica	Fílmica	Fílmica	Película	Programado según demanda

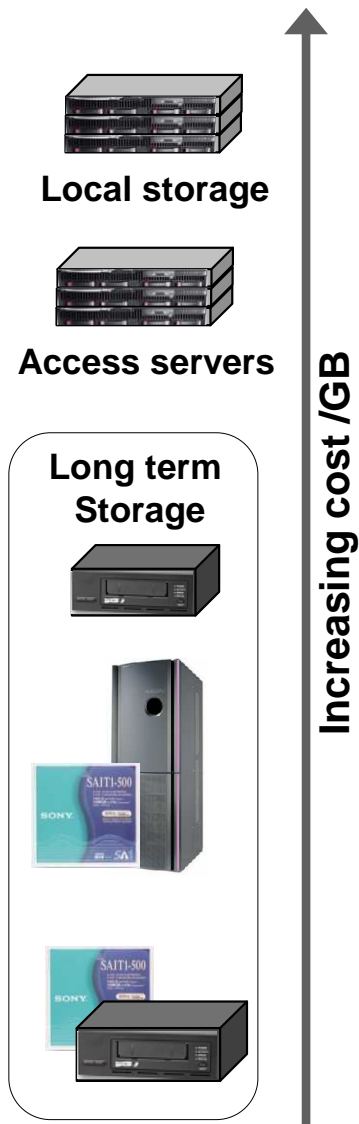
**Reglas y procedimientos son VITALES**  
**Demasiados procesos paralelos**  
**Muchos procesos son automáticos**  
**Demasiados archivos, versiones, variables**  
**REGLAS FIJAS Y REVISADAS CONSTANTEMENTE**  
**DEBEN APLICARSE SIEMPRE**  
**DEBEN IMPLEMENTARSE CHEQUEOS Y CONTROLES**  
**AUTOMÁTICAMENTE TAMBIÉN, SI ES POSIBLE**





**Metadatos técnicos y administrativos son VITALES!**  
**Es crítico que sean recogidos**  
**IMMEDIATAMENTE**  
**DESDE EL PRINCIPIO**  
**AUTOMATICAMENTE SI ES POSIBLE**





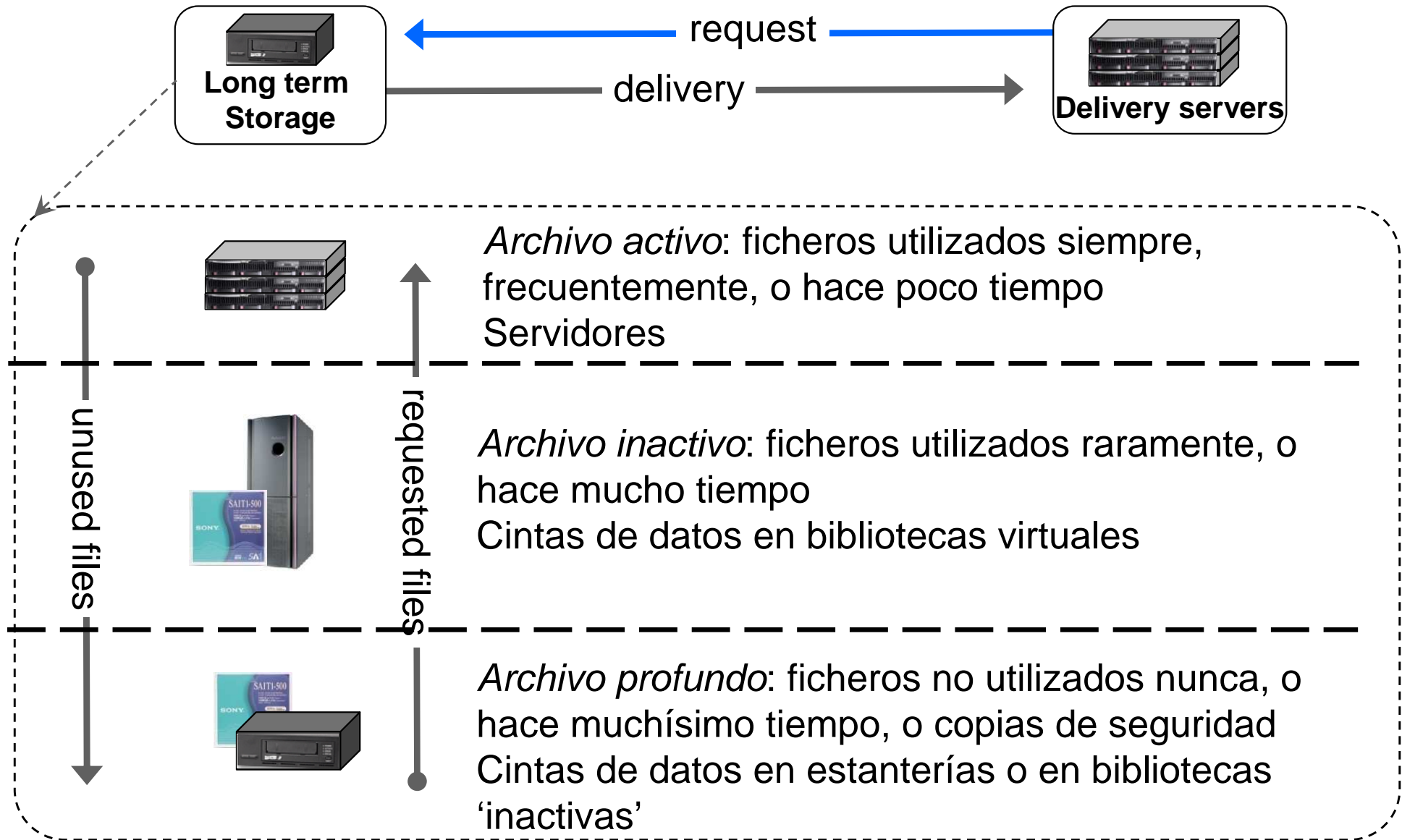
## Tipos de Almacenamiento

**Almacenamiento local:** servidores de máximo rendimiento – acceso rápido, alta velocidad, control absoluto, tiene que gestionar ficheros enormes, capacidad depende de la productividad y del flujo de trabajo

**Acceso:** generalmente servidores web – acceso rápido, velocidad media, buen control, ficheros de dimensión media o pequeña, varios streams paralelos, la capacidad depende del material online

**Almacenamiento a largo plazo:** generalmente cintas de datos (bibliotecas virtuales o estanterías) – acceso más lento, velocidad limitada, buen control, migración fácil en bibliotecas (pero NO en estanterías), ficheros de cualquier dimensión, capacidad ilimitada

# Almacenamiento Jerárquico Multi-Nivel







## Funciones primarias de un sistema de Gestión del Almacenamiento:

- Ingesta
- Mantener la sincronización con las bases de datos
- Creación automática de metadatos (ej: cuando se hacen copias o los ficheros migran o son transferidos)
- Búsqueda y recuperación
- Control del acceso: 'sólo acceso', 'acceso y modificar', 'acceso-modificar-suprimir', etc.
- Entrega a los servidores de acceso (webservers, VTRs, etc. **Si es posible**, controlar el proceso y las máquinas)
- Gestión de los sistemas de almacenamiento (discos, cintas, etc.
- Control de las versiones y de las copias: ej: gestión de diferentes versiones del mismo título y de diferentes copias idénticas en **lugares distintos**
- Mover ficheros de un almacén a otro (ej: dentro de un almacenamiento jerárquico, o de/a servidores de acceso)
- Control de calidad y chequeo (cuando se copian o mueven ficheros)
- Migración automática de soportes y formatos
- Eliminación de ficheros

## Funciones primarias de un sistema de Gestión del Almacenamiento:

- Ingesta
- Mantener la sincronización con las bases de datos

**En un primer momento muchas de estas funciones se realizan manualmente, pero NO cuando las colecciones digitales crecen en tamaño y complejidad (cantidad de datos, número de archivos). Debe proyectarse una sala para su crecimiento desde el primer momento.**



Long term Storage



- diferentes versiones del mismo título y de diferentes copias idénticas en **lugares distintos**
- Mover ficheros de un almacén a otro (ej: dentro de un almacenamiento jerárquico, o de/a servidores de acceso)
  - Control de calidad y chequeo (cuando se copian o mueven ficheros)
  - Migración automática de soportes y formatos
  - Eliminación de ficheros

## Primary functions of a Storage Mgmt system:

- Ingest
- Maintain sync with database

Y...

**Todas estas funciones descansan casi por completo en *Metadatos técnicos, administrativos y descriptivos!***

- Move files from one subsystem to another (eg within hierarchical storage, or to/from delivery servers)
- Check and quality control (when files are actively copied or not)
- Automatic Media (and Format) migration
- File de-accessioning



Don Loviz