



Clase 4

Administración

Tareas cotidianas en el repositorio

Análisis documental | Catalogación

Es un conjunto de operaciones destinadas a **representar** la forma y el **contenido** de un documento según **normas estandarizadas**. Es el proceso que permite el reconocimiento de un documento a partir de una **descripción unívoca** y sin ambigüedades proporcionando los elementos necesarios para su identificación.



Análisis documental | Catalogación

Objetivos

- Identificar los documentos de forma inequívoca
- Dar uniformidad a la información
- Organizar y preservar el material alojado en una institución
- Facilitar la ubicación, el acceso y la recuperación
- Permitir el intercambio con otras instituciones



Normalización

La normalización persigue **regularizar objetos** que, en la práctica, no suelen ser necesariamente uniformes, con el fin de **facilitar las operaciones** documentales, disminuyendo costes, mejorando la **calidad**, **acelerando procesos** y posibilitando intercambios.

La finalidad de usar normas técnicas en diferentes Unidades de Información y servicios de información es proveer **compatibilidad** e **interoperabilidad** entre equipos, datos, información, políticas y procedimientos.



Lista de normas

- Reglas de catalogación
 - [ISBD](#)
 - [AACR2](#)
 - [RDA](#)
 - [FRBR](#)

 - **Números normalizados**
 - ISSN
 - ISBN

 - **Sistemas de clasificación**
 - [Dewey](#)
 - [CDU](#)
- **Lenguajes documentales**
 - Tesauro
 - Lista de epígrafes

 - **Lenguajes naturales**
 - Palabras clave

 - **Normas ISO**



Análisis documental

Catalogación, descripción bibliográfica: describe el documento para su **identificación física**. Análisis de la **forma**: título, autor, páginas, editorial, idioma, etc.

Indización, clasificación: se escogen los conceptos más adecuados para representar el **contenido** de un documento (de qué trata) y la **traducción** de estos conceptos a **términos** extraídos de un **lenguaje documental**. Se pretende desarrollar un sistema que permita la búsqueda de los documentos por los temas que tratan.

Resumen: breve descripción del contenido de un documento

Indización, clasificación (análisis de contenido)

Consiste en:

- La búsqueda, análisis y selección de los conceptos presentes en el documento para **describir su contenido**, con miras a su almacenamiento y recuperación.
- La **traducción** de esos conceptos para pasar del lenguaje natural del autor (el que utiliza el discurso científico, técnico o literario) al **lenguaje documental o controlado**, para evitar metáforas, sinonimias y homonimias.

Indización, clasificación (**Herramientas**)

Tesauros: listas de términos (palabra o conjunto de palabras) llamados “**descriptores**” que guardan entre sí relaciones semánticas de **jerarquía, asociación y equivalencia**.

Listas de encabezamientos de materias: listas de términos (palabra o conjunto de palabras) **ordenadas alfabéticamente**. A estos términos se les denomina “encabezamientos de materia” y bajo cada uno de ellos se indican los términos **sinónimos, genéricos o específicos** con los cuales se relaciona. A partir de la determinación del **tema principal** de un documento se asigna el encabezamiento de materia que le corresponde para ese tema.

Indización, clasificación (otros recursos)

Palabras clave

- Vocablos extraídos del **lenguaje natural**
- Se extraen del título o del contenido del documento y carecen de sistematización.

Lista de descriptores libres: listas de términos (palabra o conjunto de palabras) que elabora el **bibliotecario** según las necesidades de su biblioteca y de sus usuarios.

Análisis de contenido - Tesauro Eurovoc (SOC)



Europa > Página inicial de EuroVoc > Campos temáticos & MT > 0406 marco político

Lengua del contenido:
 (es) Español

Búsqueda

Búsqueda avanzada

Examinar
 Examinar la presentación temática

Descargar
 Por campo temático
 Presentación alfabética permutada
 Lista multilingüe
 Índice alfabético
 SKOS/XML

Sus propuestas
 Contribuir
 Nuevos conceptos aprobados

0406 marco político

Estado
 NT1 Estado confederal
 NT1 Estado confesional
 NT2 Estado islámico
 RT Derecho musulmán [1206]
 NT1 Estado de Derecho
 RT primacía del Derecho [1206]
 NT1 Estado del bienestar
 RT bienestar social [2821]
 RT política de intervención [1606]
 NT1 Estado federal
 RT federalismo [0406]
 NT2 Estado federado
 NT1 Estado laico
 RT laicismo [0406]
 NT1 Estado multiétnico
 RT grupo étnico [2821]
 NT1 Estado regional
 RT Parlamento regional [0421]
 RT regionalismo [0406]
 NT1 Estado unitario
 NT1 símbolo del Estado
 RT bandera [2831]

Análisis de contenido - Sistemas de clasificación

Un libro sobre “Edificios e instalaciones oficiales de enseñanza media”

<i>Lenguaje documental usado</i>	<i>Traducción al lenguaje documental</i>
Sistema de clasificación decimal	371.6
Lista de encabezamientos de materias	ARQUITECTURA ESCOLAR
Tesauro	ESCUELAS ESPACIOS EDUCATIVOS PLANIFICACION DE LAS INSTALACIONES NORMAS DE CONSTRUCCION DISEÑO ARQUITECTONICO
Lista de términos libres	ESCUELAS-RANCHO



Resumen o *abstract*

Representación abreviada del contenido de un documento sin interpretación ni crítica.

Los objetivos del resumen son:

- **Informar al usuario** sobre el **contenido** preciso de los documentos. Proporciona al usuario los **elementos decisorios** sobre la conveniencia de consultar el original íntegro o de desechar su selección por no ser de interés. Determina la pertinencia del documento.
- Es utilizable para la **recuperación automatizada**
- Para las personas con un **interés marginal** por el tema les es suficiente con leer el resumen.

Normalización de nombres de autores

La falta de normalización de los nombres de los investigadores y de sus centros en las publicaciones científicas y en las principales bases de datos bibliográficas es un hecho conocido, que disminuye la visibilidad de los autores y de sus centros a nivel nacional e internacional y dificulta la recuperación de las publicaciones y de las citas por ellas recibidas.

VIAF (Fichero de Autoridades Virtual Internacional) es un proyecto conjunto de varias bibliotecas nacionales, implementado y alojado por OCLC (Online Computer Library Center) <https://viaf.org/>

IRALIS (International Registry of Authors-Links to Identify Scientists). Es un sistema de estandarización de las firmas de los autores científicos. <http://www.iralis.org/es>



Normalización de nombres de autores

ORCID (Open Researcher and Contributor ID). Sistema internacional para asignar una identificación de autor mediante una URL para facilitar la recuperación de información bibliográfica sobre sus trabajos en diferentes recursos. Es un proyecto apoyado por la mayoría de las grandes editoriales a nivel mundial.

Researcher ID. Permite generar un código identificador de un autor de forma inequívoca para facilitar las búsquedas y la creación de perfiles curriculares y datos bibliométricos como las citas recibidas.

Mendeley. Además de gestor bibliográfico, permite crear un perfil curricular a modo de red social académica donde gestionar las publicaciones de un autor.

Research Gate. Red social académica que permite mostrar y compartir publicaciones.

Google Scholar Citations. Herramienta de Google para crear un perfil profesional de publicaciones y hacer un seguimiento a las citas recibidas.



Metadatos



Metadatos

Son la información descriptiva que se aplica sobre un recurso para facilitar su organización, son datos que describen otros datos, información estructurada para describir, explicar, localizar o facilitar la obtención, uso o administración de un recurso de información.

Según la norma ISO 15489-1 del 2001 define los metadatos, en el contexto de la gestión de documentos, como: "datos que describen el contexto, contenido y estructura de los documentos, así como su gestión a lo largo del tiempo (...) Como tales, los metadatos son información estructurada o semiestructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo y dentro de un mismo dominio o dominios diferentes."

Papel de los metadatos

- Recuperación de la información
- Administración de documentos
- Gestión de derechos, autoría y propiedad intelectual
- Estado de archivo
- Control y descripción de procesos



Papel de los metadatos

- Seguridad y autenticación
- Valoración de contenidos
- Preservación y conservación
- Visibilidad de la información
- Actualización de la información



Tipos de Metadatos

Tipo	Uso	Ejemplos
Administrativo	Usados en la identificación, gestión y administración de recursos de información	Adquisición de información Derechos y reproducción Requerimientos legales para el acceso Localización de información Criterios de selección para la digitalización Control de la versión
Descriptivo	Utilizados para representar recursos de información	Registros catalográficos Proporcionar ayuda en la búsqueda Índices especializados Hiperenlazar relaciones entre recursos Anotaciones de los usuarios
Preservación	Para salvaguardar los recursos de información	Informar sobre las condiciones de uso de los recursos físicos Informar sobre las acciones llevadas a cabo para preservar versiones físicas y digitales de recursos
Técnico	Relativos a cómo funcionan los sistemas o el comportamiento de los metadatos	Documentación de hardware y software Digitalización de la información (formato, ratio de compresión...) Autenticación y datos de seguridad (encriptación, passwords, etc.) Control de tiempo de respuesta de sistemas
Uso	Relativos al nivel y tipo de uso que se hace con los recursos informativos	Información sobre versiones Reutilización del contenido del recurso

Metadatos – Dublin Core (DCES)

Dublin Core DC es un esquema de metadatos elaborado por la DCMI (Dublin Core Metadata Initiative). Actualmente es el esquema de datos más utilizado. Consta de 15 elementos básicos.

Contenido	Propiedad	Instanciación
	Intelectual	
Title	Creator	Date
Subject	Publisher	Type
Descripción	Contributor	Format
Source	Rights	Identifier
Lenguaje		
Relation		
Coverage		



Metadatos – Dublin Core cualificado (DCTERMS)

El [Dublin Core cualificado](#) es una extensión de esquema Dublin Core básico donde algunos de sus elementos están especializados:

- **Título:** dc.title ⇒ dcterms.alternative, dcterms.title
- **Relación:** dc.relation ⇒ dcterms.isVersionOf, dcterms.isPartOf, ...
- **Fecha:** dc.date ⇒ dcterms.created, dcterms.available

Otros tipos de esquema son:

- Networked Digital Library of Theses and Dissertations (ETD)
- Metadata Objects Description Schema (MODS)
- Categories for the Description of Works of Art (CDWA)
- Cataloguing Cultural Objects (CCO)



Networked Digital Library of Theses and Dissertations (ETD)

- Universidad de Utah State
 - Versiones electrónicas de tesis o disertaciones
 - Metadatos utilizados en SEDICI
-
- `thesis.degree.grantor` (Institución garante)
 - `thesis.degree.name` (Grado alcanzado)



Metadata Objects Description Schema (MODS)

- Desarrollado por la Oficina de Desarrollo de Redes y Normas MARC de la Biblioteca del Congreso (LC)
- MODS es un conjunto de metadatos que tiene el mismo objetivo que DC, pero que busca superar los problemas y limitaciones detectados en la aplicación de DC. Para ello, toma como punto de partida el formato MARC, del que selecciona diferentes En SEDICI se utilizan:
 - **mods.location** (URL de acceso al catalogo de la biblioteca, o información de acceso al recurso físico)
 - **mods.origenInfo.place** (Institución donde se creó el documento)
 - **mods.recordInfo.recordContentSource** (Registra el nombre de la entidad y/o sitio web desde donde se obtuvo la información del registro)

Categories for the Description of Works of Art (CDWA)

Directrices para la documentación de los bienes patrimoniales desarrolladas por un grupo de trabajo dirigido por el Getty y la College Art Association of America (CAA),

Cataloguing Cultural Objects (CCO)

Manual para documentar arte, arquitectura y objetos culturales desarrollado y diseminado por el Visual Resources Association (VRA).

Metadatos - Esquema de metadatos SEDICI

SEDICI utiliza un perfil de metadatos propio: dichos metadatos están traducidos al inglés y al portugués, y el sistema los transforma al formato [Dublin Core](#) para que sean accedidos por otros repositorios vía [OAI-PMH](#).

Ej.: [tesis](#)

dc.format.extent	110 p.	es
dc.language	es	es
dc.title	Familias y estrategias de crianza	es
dc.type	Tesis	es
sedici.title.subtitle	Prácticas de tránsito de niños y niñas por diversos grupos de crianza en sectores populares de Viedma	es
sedici.creator.person	Belmartino, María Belén 🇺🇦	es
sedici.subject.materias	Trabajo Social 🇺🇦	es
sedici.subject.decs	Familia 🇺🇦	es
sedici.subject.decs	Crianza del Niño 🇺🇦	es
sedici.subject.decs	Población Suburbana 🇺🇦	es
sedici.description.fulltext	true	es
mods.originInfo.place	Facultad de Trabajo Social 🇺🇦	es
sedici.subtype	Tesis de maestría	es
sedici.rights.license	Creative Commons Atribución-NoComercial-CompartirIgual 2.5 Argentina (CC BY-NC-SA 2.5)	
sedici.rights.uri	http://creativecommons.org/licenses/by-nc-sa/2.5/ar/	
sedici.contributor.director	Colangelo, María Adelaida 🇺🇦	es
sedici.contributor.juror	Colangelo, María Adelaida	es
sedici.contributor.juror	Fuentes, Pilar	es
sedici.contributor.juror	Velurtas, Marcela	es
thesis.degree.name	Magister en Trabajo Social 🇺🇦	es
thesis.degree.grantor	Facultad de Trabajo Social 🇺🇦	es
sedici.date.exposure	2012-06-04	



Tipologías



Tipologías y formatos

Artículo: Un artículo es un texto en prosa no ficcional en el que se desarrolla un tema en particular. Está escrito por un autor o varios autores, quienes llevaron adelante una investigación y escriben para publicar los resultados de investigación en una revista científica o académica.

Sonido: Recurso diseñado principalmente para ser oído. Ejemplos: formato de archivo de audio, CD-ROM y grabaciones de sonido. (AUDIO EN CIC)

Comunicación de congreso: Comunicación de congreso que se presenta en el congreso y se expone ante la audiencia. La comunicación del congreso se publica en las actas.

Fuente: Centro Argentino de Información Científica y Tecnológica [Vocabulario de tipos de recursos de información](#)



Contribución a revista: Texto (por ejemplo: artículo, columna, editorial u otra contribución) escrito para ser publicada en una revista académica o científica.

Imagen fija: Representación visual estática. Este tipo de imágenes incluye diagramas, dibujos, diseños gráficos, planos, mapas, fotografías e impresiones.

Informe: Se denomina informe al registro de hallazgos de una investigación terminada o en progreso, o de otros hallazgos técnicos. Por lo general, está identificado por un número de informe y a veces también por un número de subsidio asociado a la agencia de financiamiento.

Libro: Una publicación no seriada que está completa en un volumen o en un número definido de volúmenes. En su formato físico, un libro es una serie de hojas de papel, pergamino u otro material encuadernadas en uno de sus extremos con tapas. Los libros se suelen identificar con un ISBN.

Capítulo de libro: Un capítulo o sección de un libro, por lo general con un título o número de capítulo que lo identifica.

Publicación Periódica: Publicación seriada con su propio título identificador, caracterizada por una variedad de contenidos y contribuciones (por ejemplo, artículos, editoriales, reseñas, columnas), que se publica con una periodicidad regular.

Reseña: Una reseña del trabajo de un tercero.

Tesis: una tesis o disertación es un documento presentado para postularse a un título académico o a un cargo profesional, en la que el autor presenta un tema de investigación y los resultados. Pueden ser de grado, maestría o doctorado.



Formatos de Archivos



Formatos Digitales

Un formato digital es un sistema particular de codificación de la información para su almacenamiento o tratamiento en un soporte informático o digital.

Los formatos más conocidos son el *.doc* para texto, *.tiff* y *.jpeg* para imágenes, *.mp3* y *.wav* para audio, *.avi*, *.mpg* para video, etc.



Los objetivos perseguidos por un formato específico pueden ser diversos:

- Almacenar un solo tipo de contenido plano sin ninguna codificación adicional. Ejemplo: .txt
- Incorporar especificaciones para codificar la información (principalmente para su compresión, transmisión o cifrado) Ejemplo: .pdf
- Combinar y sincronizar varios tipos de contenido en un solo archivo. Ejemplo: los archivos .mpeg o .AVI que incluyen pistas de audio, vídeo, subtítulos, metadatos, etc.



Cuestiones a tener en cuenta

- **¿Su uso está generalizado?** ¿Existen varios programas para leer este formato? ¿Está utilizado por otras instituciones como formato de preservación?
- **¿Está abierto?** El uso del formato no debe ser regido por patentes.
- **¿Está documentado?** ¿Fue publicada su documentación (un formato puede estar documentado sin estar necesariamente abierto)? ¿Fue normalizado por instituciones como W3C o ISO? Esta documentación permitirá construir nuevos programas para leer el formato si sus vendedores ya no se encargan de su mantenimiento.
- **¿Existen programas para validarlo y caracterizarlo?**



Dificultades para la elección de Formatos

- Cuando uno elige el formato más adecuado al contenido, debe valorar los riesgos de **obsolescencia** que corre el formato.
- También se debe tomar en cuenta si las políticas del repositorio o la legislación de la institución impondrán que sean aceptados todos los datos, sea cual sea su formato.
- La finalidad del archivo digital y su ubicación en la cadena de valor: formato inicial de **creación**, formato intermedio para **edición**, formato final para preservación o **difusión**. A veces el formato mostrado al público no es el mismo en el que se preserva.

Niveles de Riesgo de un Formato (BnF)

Nivel	Descripción
Almacenado	Formato no identificado para el cual sólo se asegura la preservación de la cadena de bits.
Identificado	Formato identificado por una herramienta de identificación para el cual no fue definido ninguna estrategia de preservación.



Nivel	Descripción
Conocido	Formato para el cual la BnF tiene por lo menos una herramienta de referencia, realiza una vigilancia tecnológica y define una estrategia de preservación, sea emulación o migración a un formato controlado.
Controlado	Formato documentado por el cual la BnF tiene herramientas de referencia, realiza una vigilancia tecnológica e impone requisitos a los productores.



Factores de sostenibilidad a tener en cuenta en la selección de los formatos

Divulgación/Transparencia: hace mención al grado en el que las especificaciones técnicas de los formatos son accesibles (es decir son de código abierto).

Apertura: se refiere a la dependencia de un formato determinado a una patente. En el caso de las entidades dedicadas a la preservación y/o difusión del patrimonio cultural, en general, se tiende a utilizar formatos abiertos, es decir formatos de archivo que no se hallan sujetos a patentes o derechos de autor. Ejemplo: Empresas como Apple, Amazon, Microsoft conforme con criterios comerciales y no necesariamente de preservación, pueden imponer formatos propietarios, es decir, protegidos por patentes o derechos de autor.

Dependencia/Interoperabilidad: se trata del grado de dependencia de un formato determinado respecto a un hardware, a un software o a un sistema operativo específico.

Estabilidad/Compatibilidad: grado en el que un formato mantiene su funcionalidad e integridad con versiones anteriores o posteriores.

Aceptación: hace referencia al grado de utilización de los formatos por parte de los creadores, distribuidores y usuarios de los recursos.

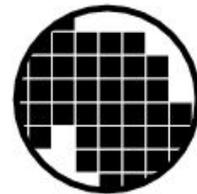
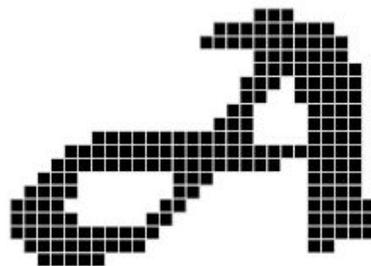
Estandarización: adecuación formal a los procesos o especificaciones establecidos por un organismo de normalización (ISO, NISO, W3C...) con el objetivo de garantizar la calidad de los archivos y su interoperabilidad.

Mecanismos de protección técnica: los mecanismos de protección técnica como por ejemplo el cifrado, utilizados habitualmente para proteger la propiedad intelectual, no deben dificultar la recuperación de datos, migración de los contenidos o su adaptación a nuevas necesidades derivadas de la evolución tecnológica.

Formatos de Imagen

Existen dos tipos básicos de formatos de imagen:

- Gráficos o imágenes pixeladas o de **mapa de bits** (raster images): las imágenes están representadas por píxeles. Es el tipo de formato que se utiliza comúnmente para captar imágenes con cámaras digitales y escáneres.
- imágenes **vectoriales**: están definidas matemáticamente y utilizan formas geométricas para representar las imágenes. En este caso la información que almacenan no hace referencia a los píxeles. Se utilizan principalmente en actividades como el diseño gráfico, ingeniería, etc. El tamaño de los ficheros es muy pequeño.



Las dos características básicas que determinan la calidad de la imagen, principalmente en las imágenes pixeladas, son la **resolución** y la **profundidad de color** o profundidad del bit.

La **resolución** es la cantidad de puntos que componen una imagen, es decir, la capacidad de distinguir los detalles espaciales más finos. Habitualmente se expresa en píxeles por pulgada (ppp o ppi en inglés). Lo importante es que a mayor número de píxeles, mayor calidad de la imagen y mayor peso del archivo.

La **profundidad de color** o **profundidad del bit**, se refiere al número de bits utilizados para describir el color de cada píxel. Cuanto mayor sea la profundidad de bits mayor será la cantidad de tonos representados

Número de colores visibles por bit

Profundidad de Color	Nº de colores visibles
1 bit (monocromo)	2 tonos
4 bit	16 tonos
8 bit (indexed color)	256 tonos
24 bit (true color)	16.777.216 tonos

Formatos de Imagen Recomendados (IFLA)

Profundidad de Color	Nº de colores visibles
Preservación documentos impresos y manuscritos y otros materiales especiales (fotos, planos, grabados, mapas...)	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises 256 gamas / 16,7 millones colores
Preservación prensa histórica	Formato TIFF sin compresión / Resolución mínima 400 ppp / Color: Escala grises (256 gamas)
Preservación prensa moderna	Formato TIFF sin compresión / Resolución mínima 300 ppp / Color: Escala grises (256 gamas)
Difusión	archivos con compresión JPEG

* En la digitalización de imágenes, el TIFF sigue siendo el formato más utilizado en la preservación de documentos, aunque el JPEG 2000 está comenzando a ser usado siempre que la compresión sea sin pérdida.



Formatos de Audio



Se distinguen dos categorías elementales de formatos:

- Por un lado, aquellos que recogen en forma de datos el sonido creado directamente a través de dispositivos electrónicos como ordenadores, sintetizadores, etc. para que después sean interpretados por el software utilizado para su reproducción.
- Por otro, los formatos más comunes y que interesan en este caso, los utilizados en las grabaciones musicales, de voz o de cualquier otro tipo, como por ejemplo el WAV, que almacena muestras de la onda de sonido para su posterior reproducción (formatos waveform).



Formatos de Audio

En contenidos de audio son dos las características principales a considerar:

- **Frecuencia de muestreo:** número de muestras que se toman de la onda de sonido en un período de tiempo, se expresa en Khz.
- **Profundidad de bits:** número de bits utilizados para representar la onda sonora.

A mayor frecuencia de muestreo y profundidad de bits, mayor calidad del sonido. Por ejemplo, para obtener una calidad de CD de audio, la grabación se realiza a 44,1 Khz y 16 bits. En cambio, cuando sólo se va a difundir, se eliminan espacios vacíos y algunas frecuencias del archivo.



Formatos de Audio Recomendados

Recomendaciones IASA (International Association of Sound and Audiovisual Archives) para preservación

Formato WAVE y BWF con fines archivísticos

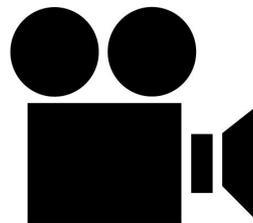
Calidad recomendable: 96 kHz / 24 bit

Calidad mínima: 48 kHz / 24 bit

Las mismas especificaciones para las grabaciones de voz



Formatos de Video



Los formatos de vídeo son en general, **formatos contenedores** o formatos de compresión de información audiovisual ya que contienen varios tipos de información: pistas de vídeo, pistas audio, en ocasiones texto o incluso imágenes incrustadas para los subtítulos y los metadatos necesarios para su sincronización.

En el caso de los contenidos audiovisuales resulta más complejo establecer unas directrices de uso común para la preservación por los diferentes tipos de información (audio, imagen...) y los productos y canales de difusión específicos para los que están diseñados: grabación, edición, cine, DVD, internet.

Resolución de la imagen: definida por el número de píxeles («puntos») que componen la imagen.

Velocidad de la imagen (frame rate): número de imágenes estáticas o fotogramas por segundo, se expresa en fps (frames per second). Existen distintos estándares para cine (24 fps) y los distintos sistemas de televisión, etc.

Tasa de bits (bit rate): velocidad de transferencia de los datos contenidos en el archivo audiovisual. Se expresa en bits por segundo (bps) y determina la nitidez de la imagen: a mayor velocidad más información puede contener la imagen.

Profundidad de color: al igual que en las imágenes estáticas, se refiere al número de bits utilizados para describir el color de cada píxel.

Relación de aspecto o tamaño del cuadro: proporción entre el ancho y la altura de la imagen.

Formatos de Video Recomendados

Preservación	Difusión
JPEG2000 (ISO/IEC 15444-4) (.mj2)	JPEG 2000 MPEG4 AVI MOV MP4 MKV OGG



Formatos de Texto

El formato más difundido para la publicación de textos es el formato de documento portátil o **PDF**, pero no es raro encontrar autoarchivos de materiales hechos en formatos **.doc** o **.docx** u otro tipo de formatos de texto editable como **.odt**. En estos casos siempre es recomendable la transformación del material al formato PDF. El formato PDF fue creado por Adobe, y es ahora un estándar abierto y oficial reconocido por la Organización Internacional para la Estandarización (ISO). A los fines de la preservación digital el formato recomendado es el PDF/A. El **PDF/A** es el estándar más común para los documentos de texto con formato, pero muchas entidades que ofrecen contenidos en formatos de texto electrónico en formato **EPUB**. Ambos formatos están basados en XML.

Formatos recomendados para Texto

El **PDF/A** se presenta como el estándar aceptado para la creación de documentos digitales accesibles online y susceptibles de ser impresos, tanto aquellos basados en texto como los que incluyen imágenes, gráficos, etc. que requieren de un diseño preciso.

El **EPUB** es el estándar de facto recomendado para el texto electrónico. Aunque puede soportar imágenes está más orientado a la publicación de texto, por ello no es el formato más adecuado para documentos que requieren un diseño preciso o están basados en imágenes.

Ventajas del PDF/A

PDF/A es, de hecho, un subconjunto de PDF obtenido excluyendo aquellas características superfluas para el archivado a largo plazo de forma similar a como se ha definido el subconjunto PDF/X para la impresión y artes gráficas. Además, el estándar impone una serie de requisitos a los programas para la visualización de archivos PDF/A.

Un programa de visualización que se ajuste a los requisitos debe seguir ciertas reglas incluyendo la conformidad con las directrices en cuanto a la gestión de color, el uso de fuentes integradas a la hora de la visualización, o la posibilidad de realizar anotaciones por parte del usuario.

En síntesis:

- Para la preservación se utilizan en mayor medida los formatos no propietarios, reconocidos como estándares. Cuando se trata de difundir existe una mayor flexibilidad.
- Los formatos presentados como válidos para la preservación, también lo son para la difusión de los contenidos aunque se suele aplicar algún tipo de compresión o se reduce de calidad en pos de su funcionalidad.
- La principal divergencia encontrada en los formatos utilizados en la industria y en el ámbito de la preservación, reside en la utilización de medios de protección técnica.

Formatos Sugeridos

Tipo de contenido	Preservación	Difusión
IMAGEN	TIFF (sin compresión) JPEG2000 (sin pérdida)	JPEG JPEG2000 (con pérdida) TIFF (con compresión)
TEXTO	EPUB (libro electrónico) PDF/A1 (ISO 19005-1)	EPUB PDF/A PDF HTML XHTML Open office Texto plano (ISO 8859-1)
AUDIO	BWF WAVE	MP3
VIDEO	JPEG2000 (ISO/IEC 15444-4) (.mj2)	JPEG 2000 MPEG4 AVI MOV

Fuente: "Formatos de Difusión y Formatos de Preservación de Contenidos Digitales" Kultura 2.0-Cuadernos de Formación Gobierno Vasco - Departamento de Educación, Política y Cultura



Comunidades y Colecciones



Comunidades y colecciones

- Las comunidades pueden contener subcomunidades o colecciones, pero no ítems
- Las colecciones sólo pueden contener ítems
- Un ítem pertenece a una sola colección, pero puede estar asociado a otras colecciones

Servicios de un repositorio digital

- Búsqueda y recuperación
- Exploración
- Copyright
- Autoarchivo
- Servicios a otros sistemas



Búsqueda y recuperación



- Un repositorio digital puede alojar cientos, miles o millones de recursos
- Es necesario proveer a los usuarios de mecanismos para buscar y recuperar estos recursos.
- Los usuarios pueden estar buscando un recurso específico y único, o pueden requerir recursos con alguna característica en común (por ejemplo, artículos que traten sobre determinada área del conocimiento).
- A veces, los usuarios no saben bien que están buscando; suelen refinar los criterios de búsqueda una y otra vez hasta que localizan los recursos.



Búsqueda y recuperación

Un repositorio tiene que proveer un servicio de búsqueda simple, que permita ingresar algunos términos de búsqueda y retorne un conjunto de recursos como resultado.

También debe proveer una búsqueda avanzada, que permita parametrizar los criterios de búsqueda y acotar así el conjunto resultante: por fecha de publicación de los recursos, por tipo de recurso, por idioma, por autor...

En cualquier caso, las búsquedas deben cumplir ciertos criterios mínimos.

Búsqueda y recuperación: criterios mínimos

Simpleza: el formulario de búsqueda debe ser simple y mostrar campos de búsqueda avanzada si el usuario lo requiere. De todos modos, la búsqueda avanzada también debe permanecer simple.

Eficiencia: las búsquedas deben resolverse casi inmediatamente, en cuestión de milisegundos, o muy pocos segundos a lo sumo.

Relevancia: todos los resultados de una búsqueda tendrán un valor de relevancia. Cuanto más relevante, más arriba deberá mostrarse entre los resultados.

Búsqueda y recuperación

Filtrado: la búsqueda avanzada permite definir ciertos criterios a aplicarse durante la búsqueda

- En ocasiones, es deseable aplicar filtros una vez realizada la búsqueda.
- Para ello, es necesario definir criterios de agrupamiento de resultados, y permitir al usuario agregar o eliminar criterios.
- Una técnica muy utilizada es el faceting (faceted search, faceted navigation o faceted browsing), que permite a los usuarios explorar filtrando la información disponible en los resultados de la búsqueda.

Exploración

- Mediante la exploración, los usuarios pueden acceder a los recursos a partir de un orden preestablecido.
- Este orden puede variar de repositorio en repositorio: colecciones, temas, fechas, etc.
- La exploración permite obtener un pantallazo general del repositorio.



Trabajo Cotidiano



Administración: ingesta

¿Cómo llegan los materiales?

- 👉 **Autoarchivo**
- 👉 **Carga mediada**
- 👉 **Sword (OJS, interoperabilidad)**
- 👉 **Otros repositorios y portales**
- 👉 **Sitios específicos de las revistas/congresos**
- 👉 **Convenios con otras instituciones**
- 👉 **Personalmente**
- 👉 **Otros**



Administración: ingesta

Autoarchivo

Los materiales son enviados por los usuarios a través de un formulario de carga sumamente sencillo. En el caso de SEDICI es usado mayoritariamente por los tesisistas de **posgrado**, quienes por **mandato (resolución UNLP 78/11)** deben depositar sus tesis para obtener su título. No obstante, también lo utilizan tesisistas de **grado** así como investigadores que ni bien publican suben sus trabajos al repositorio. También es utilizado por miembros de laboratorios e institutos de la UNLP que muestran sus materiales en el repositorio (caso [LEEM](#)).



Administración: ingesta

Pre-carga mediada

En este caso, se trata de colecciones particulares cuyos **administradores** son **externos** al repositorio. Luego de haber recibido una **capacitación** en el uso de DSpace y algunas nociones de catalogación, envían sus materiales cargándolos directamente en la colección que les corresponde y luego se completa la catalogación por parte de los administradores. Es el caso de las colecciones pertenecientes a los **museos de la UNLP, REA y Radio Universidad**.

Administración: ingesta

Sword (OJS)

Mediante el protocolo Sword se **importan** al repositorio los recursos de las revistas que se gestionan a través del **Portal de Revistas** de la UNLP. Este **plugin** permite que los administradores del repositorio reciban tanto los archivos como los **metadatos básicos** de los recursos y sólo deban **normalizarlos**, y en algunos casos agregarles otros metadatos que no contempla OJS. Este plugin no se está utilizando en la actualidad debido a la última actualización de OJS.

Administración: ingesta

Otros repositorios y portales UNLP

Periódicamente, los administradores del repositorio recorren sitios de la UNLP como **Memoria Académica**, **Revistas de la FAHCE** o **Libros de la FAHCE** para actualizar las colecciones correspondientes. En este caso la carga se realiza en forma manual.

Administración: ingesta

Sitios específicos de revistas/congresos u otros

Del mismo modo, se recorren periódicamente los sitios de las revistas, congresos, unidades académicas y otros dentro del ámbito de la UNLP para actualizar las colecciones correspondientes. Ejemplo: **Papel Cosido** (Bellas Artes).

Administración: ingesta

Convenios con otras instituciones

Otras instituciones, vinculadas de un modo u otro a la UNLP, alojan sus contenidos en el repositorio debido a su impacto y visibilidad. De este modo, el repositorio obtiene otra vía de ingesta de material, como las colecciones pertenecientes a la **RedUNCI** o la **Academia Nacional de Agronomía y Veterinaria**.

Administración: ingesta

Personalmente

Los usuarios pueden concurrir al repositorio en forma personal y llevar sus materiales en un pen-drive o CD, siempre que su filiación corresponda a la universidad como alumnos, docentes, becarios, no docentes o investigadores.

También se reciben materiales por e-mail, aunque no es la vía más aconsejable (se recomienda siempre utilizar el autoarchivo).

Administración: ingesta

Otros

El repositorio posee en muchos casos los datos bibliográficos de un recurso pero no su texto completo. En ese caso, siempre que sea posible, se procura **contactar al autor** y solicitarle el texto para completar esos registros. En otros casos, cuando se trata de materiales más antiguos, de los que se posee su ubicación en alguna **biblioteca** de la UNLP, se solicita a la administración de la biblioteca que se facilite el ejemplar para su **digitalización** y posterior puesta en línea (ejemplo, tesis [Patricia Zalba](#)).



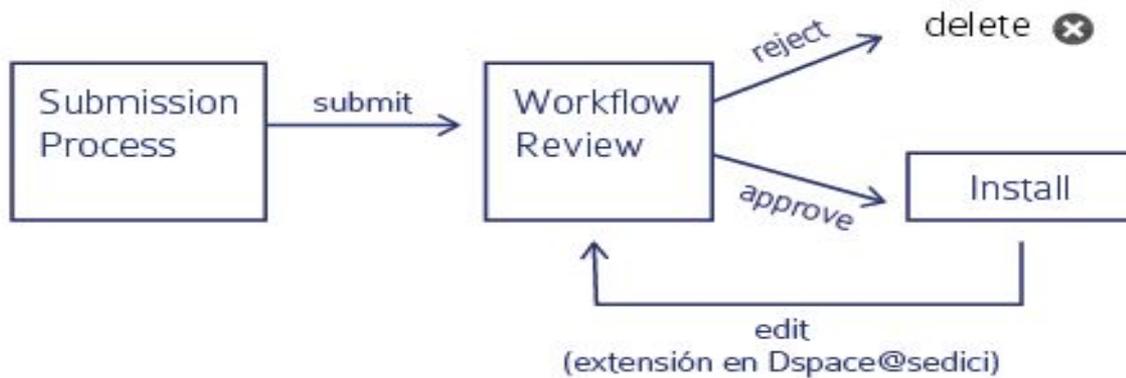


Autoarchivo

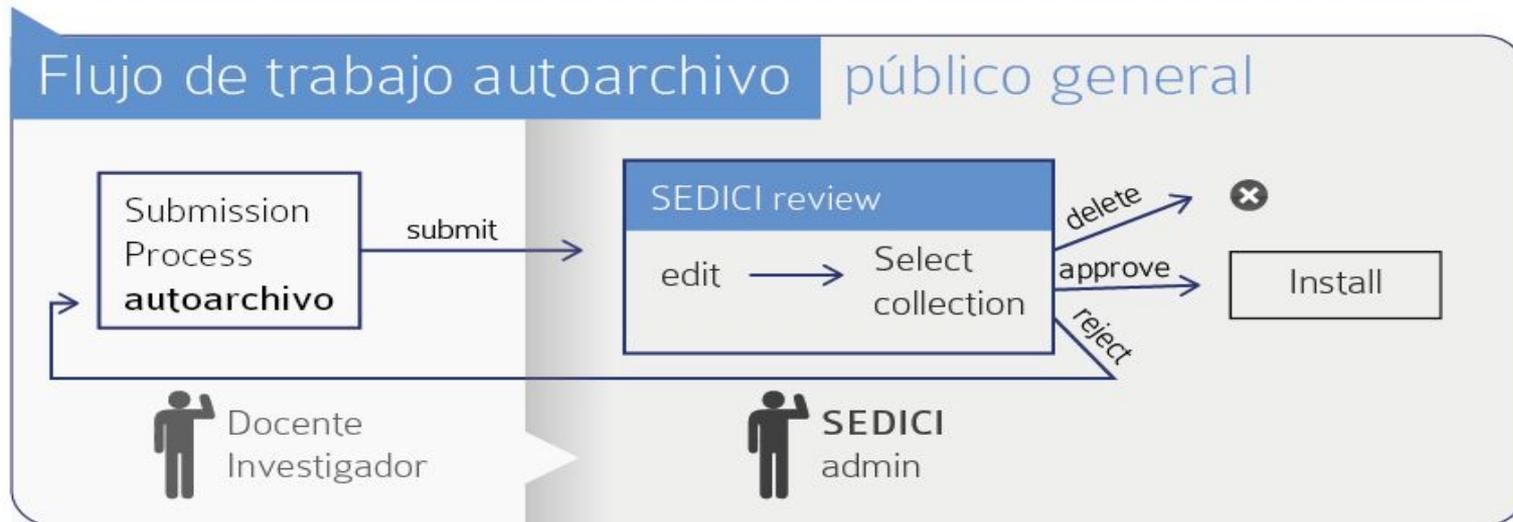


SEDICI | Flujo de trabajo

Flujo de trabajo general



SEDICI | Flujo de trabajo

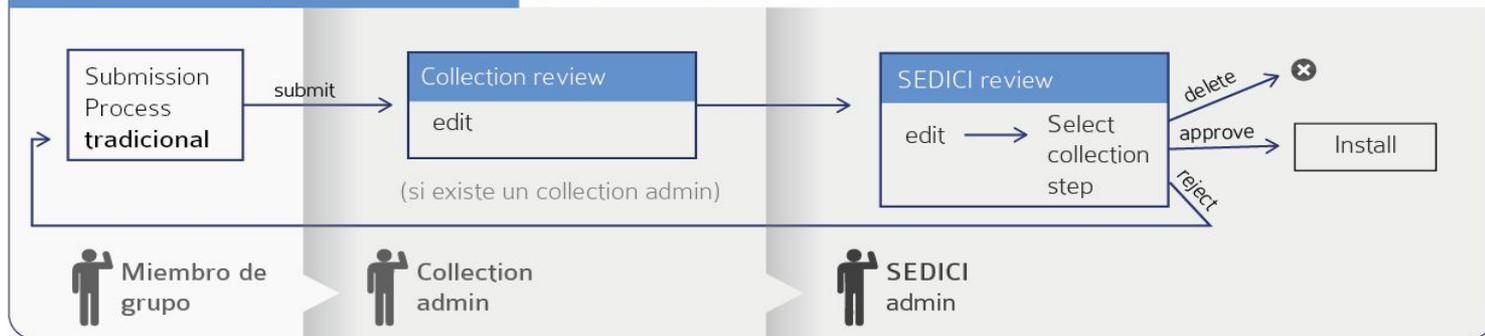


Autoarchivo



SEDICI | Flujo de trabajo

Flujo de trabajo tradicional grupos específicos



Pre-carga



SEDICI | Flujo de trabajo diario

Carga de documentos



[Analía Pinto](#)

Salir

[Editar perfil](#)

[Mis documentos](#)

Cargas Pendientes

Envíos pendientes. Podría **Nuevo envío**.

Título	Colección	Remitente
No hay envíos incompletos.		

Perfil de administrador

Tareas del flujo de trabajo

Estas tareas son ítems que están esperando aprobación antes de ser añadidas al repositorio. Hay dos colas de tareas, una para tareas que ha aceptado y otra para tareas que todavía no han sido asumidas por nadie.

Sus tareas

	Tarea	Ítem	Colección	Remitente
<input type="checkbox"/>	Revisión SEDICI	Guía de buenas prácticas para la administración de ...	Libros	Analía

Devolver las tareas seleccionadas a la cola

Tareas en cola del administrador

Tareas en cola

Tarea	Ítem	Remitente
Autoarchivo (2)		
Asumir las tareas seleccionadas		



SEDICI| Flujo de trabajo diario

Tareas en cola

Tarea	Ítem	Remitente
Autoarchivo (2)		
Asumir las tareas seleccionadas		

Tareas en cola general



SEDICI | Supervisión

Las tareas de supervisión dentro de un repositorio incluyen:

- 👉 **Distribución** de las tareas a realizar entre los administradores. Esto supone desde la correcta recepción del material por las diferentes vías (autoarchivo, recolecciones desde distintos portales dentro y fuera de la universidad, envíos por e-mail, digitalizaciones propias, convenios con terceros) hasta la asignación de qué material va a cargar cada administrador, de acuerdo a su experiencia y habilidades.
- 👉 **Revisión** exhaustiva de registros cargados en el pasado, especialmente del momento en que SEDICI no funcionaba sobre DSpace.

SEDICI | Supervisión

Las tareas de supervisión dentro de un repositorio incluyen:

- 👉 Revisión de **registros** como resultado de nuevas políticas adoptadas, por actualizaciones del sitio o por nuevas normalizaciones.
- 👉 Revisión constante de las cargas realizadas en la actualidad, para ejercer un estricto **control de calidad** sobre los datos y metadatos expuestos al mundo.
- 👉 **Correcciones** sobre los registros de diverso tipo.
- 👉 Tareas de revisión y corrección asignadas desde el **Sistema de Gestión de Incidencias** de SEDICI (también conocido como “trac”).

SEDICI | Comunicación

La comunicación dentro del repositorio puede dividirse en dos grandes grupos:

❑ **Comunicación interna**

- ❑ Soporte a los usuarios del sitio (consultas al mail, telefónicas y a través de **UserVoice**)
- ❑ **Consultas** de usuarios/tesistas sobre derechos de autor, embargo, autoarchivo
- ❑ Redacción de **tutoriales** para el uso apropiado del sitio (cómo realizar autoarchivo, por ejemplo)
- ❑ **Capacitación** de administradores internos/externos
- ❑ Redacción y edición de los manuales de procedimientos y de todo otro tutorial de uso interno para los administradores del repositorio

SEDICI | Comunicación

☐ Comunicación externa

- ☐ Presencia de SEDICI en las **redes sociales** (Twitter, Facebook, LinkedIn, SlideShare y otras).
- ☐ Redacción de posteos para el **blog de SEDICI**.
- ☐ Preparación de posteos para la fanpage de SEDICI
- ☐ Estrategias de **difusión y visibilidad**.
- ☐ **Publicidad** mediante folletos y boca a boca.



SEDICI | Comunicación interna

Manual de Procedimientos

Objetivos

- ❑ Procesar y diseminar la información recibida
- ❑ Actualizar permanentemente al personal profesional y administrativo
- ❑ Garantizar la uniformidad de tratamiento de las actividades periódicas
- ❑ Reducir los errores operativos
- ❑ Reducir el período de adiestramiento de los nuevos empleados
- ❑ Facilitar el mantenimiento de un buen nivel organizacional



SEDICI | Manual de procedimientos

Unidades académicas → Presidencia → PREBI-SEDICI → Cursos, presentaciones, tutoriales y otros

Manual de procedimientos del Servicio de Difusión de la Creación Intelectual (SEDICI)

Autor: Servicio de Difusión de la Creación Intelectual (SEDICI)

2017

Tipo de documento: Libro



Resumen

Este manual se organiza del siguiente modo: en la primera parte, titulada «Administración de...», detallan todas las tareas que hacen a la catalogación de los ítems, los metadatos utilizados, las... reciben los materiales, cómo se ordenan las colecciones y comunidades, y todo aquello que su... principal del repositorio, que es la puesta en línea, en forma adecuada y jerarquizada, ... académica, científica y artística de la universidad. En la segunda parte, titulada «Otras tareas...», diferentes roles que un administrador puede asumir así como otras tareas que se desarrollan e... repositorio, como la comunicación, la generación de contenidos para redes sociales y otros asp... en el apartado de «Anexos» se incluyen documentos de consulta para los administradores. E... decurso del manual se incluyen cuadros y recordatorios sobre distintos aspectos de la carg... facilitar la tarea diaria de los administradores; también se ofrecen enlaces para dar a cono... conceptos que provienen de otros campos del saber, principalmente de la informática.

Información general

Fecha de publicación: 2017

Editor: Pinto, Analía Verónica

Colaborador: Folegatto, Lucas (diseño de tapa) | De Giusti, Marisa (revisión técnica) | Villarrea... (revisión técnica) | Nusch, Carlos (revisión técnica) | Pelocche, Silvia (revisión técnica)

Idioma del documento: Español

Editorial: Servicio de Difusión de la Creación Intelectual (SEDICI)

Institución de origen: Servicio de Difusión de la Creación Intelectual (SEDICI)



SEDICI | Manual de procedimientos

El nuevo manual, publicado en 2017, mejoró aspectos del anterior e incorporó otros (nuevos metadatos, nuevos tipos documentales incorporados al repositorio en los últimos años, nuevas normalizaciones) y a diferencia del anterior está puesto en línea en formato **libro**, para que pueda ser difundido tanto entre usuarios comunes como entre administradores de otros repositorios que quieran tomarlo como modelo para los suyos.

SEDICI | Edición

Si bien esta no es un tarea específica de un repositorio ni tampoco cotidiana, en SEDICI se han realizado varias experiencias de edición de libros electrónicos. Los primeros de ellos fueron realizados en 2004 ([Notas sobre federalismo fiscal](#), de Horacio L. P. Piffano) y 2005 ([Microeconomía aplicada a la educación universitaria](#), del mismo autor), pero el proyecto más ambicioso encarado hasta el momento ha sido el del libro *Cirugía. Bases clínicas y terapéuticas*, único en su género.



SEDICI | Edición

Tanto en su primera como en su segunda edición, el libro incluye, además de texto, gráficos e imágenes, videos de diversas operaciones quirúrgicas relatadas por destacados profesionales de la cirugía argentina e internacional. El libro es uno de los materiales más consultados y descargados de SEDICI y a fines del año pasado se editó y publicó su tercera edición, revisada y ampliada.

Otros libros editados en SEDICI:

Análisis económico del derecho tributario, de Horacio L. P. Piffano

XVIII Ibero-American Science and Technology Education Consortium General Assembly (compilación de ponencias)

Anales de jAUTI 2012 (compilación de ponencias)



SEDICI | Autoarchivo

Demostración de cómo se realiza un autoarchivo por un usuario común y luego el camino que sigue el envío cuando es tomado por los administradores del repositorio.

