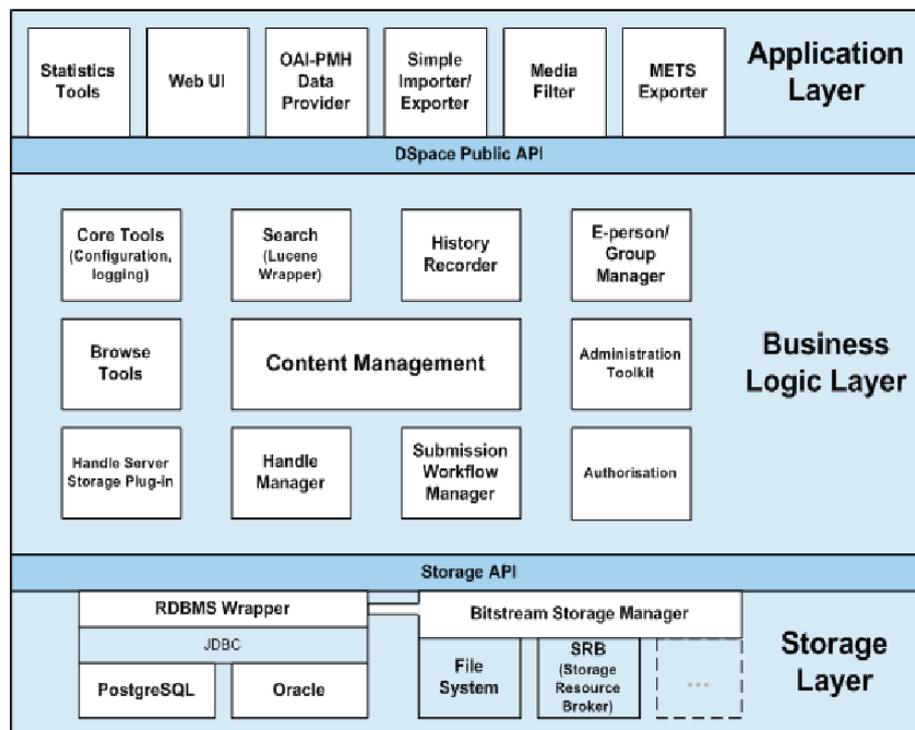


## Actividad de la Clase 2

- 1) Defina concisamente las tres capas principales que componen la arquitectura de DSpace. Describa el modelo de contenidos completo de DSpace (colecciones, comunidades, ítems, etc.).

La arquitectura DSpace se compone de 3 capas principales. Cada capa ofrece servicios a la capa superior por medio de APIs<sup>1</sup>, y utiliza los servicios de la capa inferior.



- **Capa de aplicación:** Conjunto de módulos que permiten la interacción con el mundo exterior como la aplicación web: JSUI y XMLUI, OAI-PMH Data Provider, estadísticas, importar/exportar y MediaFilter.
- **Capa de lógica de negocio:** Ofrece administración, búsqueda, exploración, gestión de usuarios y grupos, autorización, carga de

<sup>1</sup> API: *Application Programming Interface* - Interfaz de Programación de Aplicaciones

documentos, *workflow*, *handle manager*, abstracción en comunidades, colecciones e ítems.

- **Capa de almacenamiento:** Realiza la interacción con la base de datos, los ítems y sus metadatos, personas y grupos, autorización y trabajos en curso (*workflow*), índices de búsqueda y exploración. Respecto al almacenamiento de *bitstreams*:
  - Local: el almacenamiento se realiza en el sistema de archivos local al servidor en el que funciona la aplicación.
  - *Storage Resource Broker* (SRB): permite tener un sistema de archivos distribuido.

### Modelo de contenidos

- El repositorio se organiza en una o más **comunidades** de nivel base que se organizan jerárquicamente en subcomunidades descritas como espacios de trabajo. Las comunidades pueden contener subcomunidades o colecciones, pero no ítems.
- Las **colecciones** son los “estantes” dentro de las comunidades, que agrupan contenido relacionado. Las colecciones sólo pueden contener ítems.
- Los **ítems** son las obras que van en los estantes y que se pretende que el público encuentre. Un ítem pertenece a una sola colección, pero puede estar asociado a otras colecciones.
- Los **metadatos** describen al recurso.
- Los ***bitstreams*** son la representación digital del recurso. Los *Bitstreams* contienen información de validación (*checksum*), descriptiva, y de preservación (formato y nivel de soporte del formato).

- Los ítems están vinculados a sus bitstreams a través de entidades llamadas **Bundles**. Los bundles agrupan bitstreams bajo determinados criterios: ORIGINAL, THUMBNAILS, TEXT, LICENSE, CC\_LICENSE.

2) Describa al menos 2 protocolos de interoperabilidad utilizados en DSpace.

Basándose en el estándar del protocolo OAI-PMH

(<https://www.openarchives.org/OAI/openarchivesprotocol.html>), especifique los tipos de proveedores existentes y la definición de Repositorio.

Protocolos de interoperabilidad utilizados en DSpace

- **RSS y OpenSearch**: se utiliza para la sindicación de noticias y contenidos en línea. *OpenSearch* es un conjunto de tecnologías que permiten publicar los resultados de una búsqueda en un formato adecuado para la sindicación y agregación.
- **Rest**: Esta basado sobre HTTP, Muy simple y utilizado en aplicaciones web modernas, permite obtener datos o ejecutar operaciones sobre los datos, Intercambio de información en cualquier formato: XML, JSON, etc., evita las abstracciones adicionales de otros y tiene protocolos de intercambio de mensajes (ej. SOAP).
- **SWORD** (*Simple Web-service Offering Repository Deposit*) permite el depósito remoto de contenidos en el repositorio, se integra con otras aplicaciones. Existen librerías para desarrolladores en PHP, Java, Ruby, Python. El servidor SWORD expone una interfaz (ServiceDocument), a la que se le envía un documento XML, que es procesado y transformado a la representación interna del repositorio. Permite la transferencia de metadatos y binarios (archivos). Resulta particularmente útil para realizar el depósito de varios documentos a la vez.

Proveedores

**Service Provider** utiliza metadatos recolectados a través de OAI-PMH como base para construir servicios de valor agregado.

**Data Provider:** DSpace provee un módulo OAI 2.0 el cual permite al repositorio funcionar como data provider. Posee configuración flexible y adaptable basada en 4 conceptos: Contexto, filtro, transformador y mapeador.

Un repositorio digital es un medio para gestionar, almacenar, preservar, difundir y facilitar el acceso a los objetos digitales que almacena.

- 3) Describa la relación existente entre un metadato y una autoridad vinculada.  
Enumere las ventajas de utilizar metadatos controlados por autoridad.

Los metadatos son datos estructurados que describen otros datos, son datos sobre datos, que representan recursos.

Autoridad: conjunto de valores fijos identificado por una clave.

- o Registro de autoridad: información asociada con uno de los valores de la autoridad.

- o Clave de autoridad: un identificador persistente que se corresponde con el registro de autoridad.

La ubicación de las autoridades puede ser externa a DSpace, esto facilita la integración de nuevas autoridades sin modificar el código en DSpace.

En el entorno de DSpace, las autoridades sólo “existen” en los metadatos que las referencian. Cada metadato puede mantener un vínculo con una autoridad a partir del guardado de su clave y texto representativo.

Es posible conectar/vincular metadatos y autoridades a través de extensiones denominadas Choice Authority: recuperan datos desde servicios complementarios para luego guardarlos en un metadato.

- 4) Describa brevemente qué es un identificador persistente, explique las ventajas de su uso. Investigue 3 proveedores de identificadores persistentes, indique el tipo de recursos sobre los que se utilizan (artículos, imágenes, personas, etc.).

Un Identificador persistente (IP) asigna URLs fijas a recursos en línea, con el objetivo de garantizar su acceso incluso si estos cambian su URL de acceso. Las URLs pueden tener cambios en el dominio o en la ruta, cambios en el software y cambios en políticas institucionales. Los cambios de URLs deben notificarse al manejador de redirecciones, para que éste actualice sus reglas de redireccionamiento.

Proveedores de IP

**Handle:** Es un sistema abierto que permite la asignación de identificadores persistentes a los objetos digitales de Internet (artículos, revistas, imágenes, etc.), es decir, es una URL que no varía, aunque la página cambie de ubicación.

Cada handle desarrollado por la CNRI (*Corporation for National Research Initiatives*) se estructura en dos bloques:

- Prefijo: Identifica al productor (universidad, editorial, revista, etc.).
- Sufijo: Identifica a cada uno de los documentos u obras digitales (artículos, libros, capítulos, etc.)

**DOI (*Digital Object Identifier*):** Este Sistema proporciona infraestructura técnica y social para el registro y uso de identificadores persistentes interoperables para usar en las redes. Específica de contenido intelectual puede ser un artículo científico, una imagen, un libro, una canción u otro, siempre que se trate de un objeto en el

ambiente digital. El sistema DOI no cambia con el paso del tiempo, aunque el artículo sea reubicado en una dirección distinta puesto que lleva la información incorporada en forma de metadatos (Fuente: Wikipedia).

**PURL** (*persistent URL*): provee direcciones permanentes para recursos en la web. Los prefijos en común son agrupado juntos en el mismo dominio. Cada dominio tiene un solo manejador que puede agregar un nuevo PURL al dominio o hacer cambios a PURLs existentes dentro del dominio. Cada PURL tiene un *target* y un código de estado o tipo. El *target* especifica la redirección y el tipo es un código de estado de la especificación HTTP.