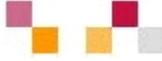




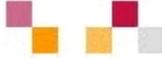
Universidad
Complutense
Madrid



Interoperatividad y estándares en la biblioteca digital....

Curso de Formación. Oposiciones Ayudantes de Biblioteca
Madrid, 18 de septiembre de 2007

Eugenio Tardón



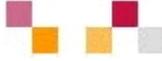
- Capacidad de dos entidades para trabajar conjuntamente con el propósito de realizar una tarea .
 - no se precisa acuerdo previo entre las instituciones que gestionan las máquinas (p.v. institucional).
 - lo que acontece cuando se proporcionan los servicios esperados (p.v. usuario).
 - capacidad de un sistema o de un producto para colaborar y trabajar conjuntamente con otros sistemas o productos sin un esfuerzo especial (p.v. de los sistemas).

- Es un cualidad fundamental: amplía visibilidad, participación y acceso a los sistemas de inf.

- Compromiso activo de asegurar que los sistemas, procedimientos y cultura de una organización se administren de manera que maximicen las oportunidades para el intercambio y la reutilización de la información



- Las bibliotecas digitales y tienen problemas para compartir e integrar datos.
- Los servicios se dan en un entorno de red heterogéneo: con múltiples tecnologías, aplicaciones, formatos de datos...
- Es preciso ocultar la heterogeneidad para dedicarnos a trabajar de forma eficaz.
- Es un problema clave: afecta al acceso a la información
- Optimiza el valor económico de la información (aumenta su consumo, visibilidad y economías de escala)



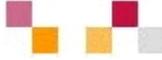
- Someterse a normas

- . Para que interoperen componentes heterogéneos, es útil acordar normas que satisfagan una cantidad limitada de homogeneidad entre ellos: las normas internacionales y fácticas.
- . El éxito o fracaso de una norma está determinado más por consideraciones sociales o comerciales que por méritos técnicos.

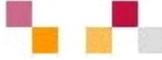
-Efectuar una mediación externa

- . Para proporcionar autonomía a los componentes es útil emplear una herramienta de mediación entre ellos. Su función básica es traducir formatos de datos y modos de interacción (Ejemplos: *gateways* de red, los *crosswalks* que convierten esquemas de metadatos)
- . Hay problemas de mediación al añadir nuevos componentes.

- Ocultar la heterogeneidad de funciones



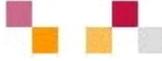
- **Normas**
- **Interfaces de usuarios**
- **Nombres e indentificadores**
- **Formatos**
- **Metadatos**
- **Protocolos de búsqueda**
- **Protocolos de red**
- **Autenticación y seguridad**



- **TCP/IP es un protocolo de comunicación para transportar datos.**
- **Un protocolo de comunicación es un conjunto de reglas y procedimientos que permiten la comunicación entre ordenadores y especifican cómo se codifican los datos en la transmisión, se detectan errores, se entrega la transferencia...**
- **Para simplificar problemas, las fases de comunicación y el software de cada fase se estratifican por capas dentro de una pila vertical de operaciones**

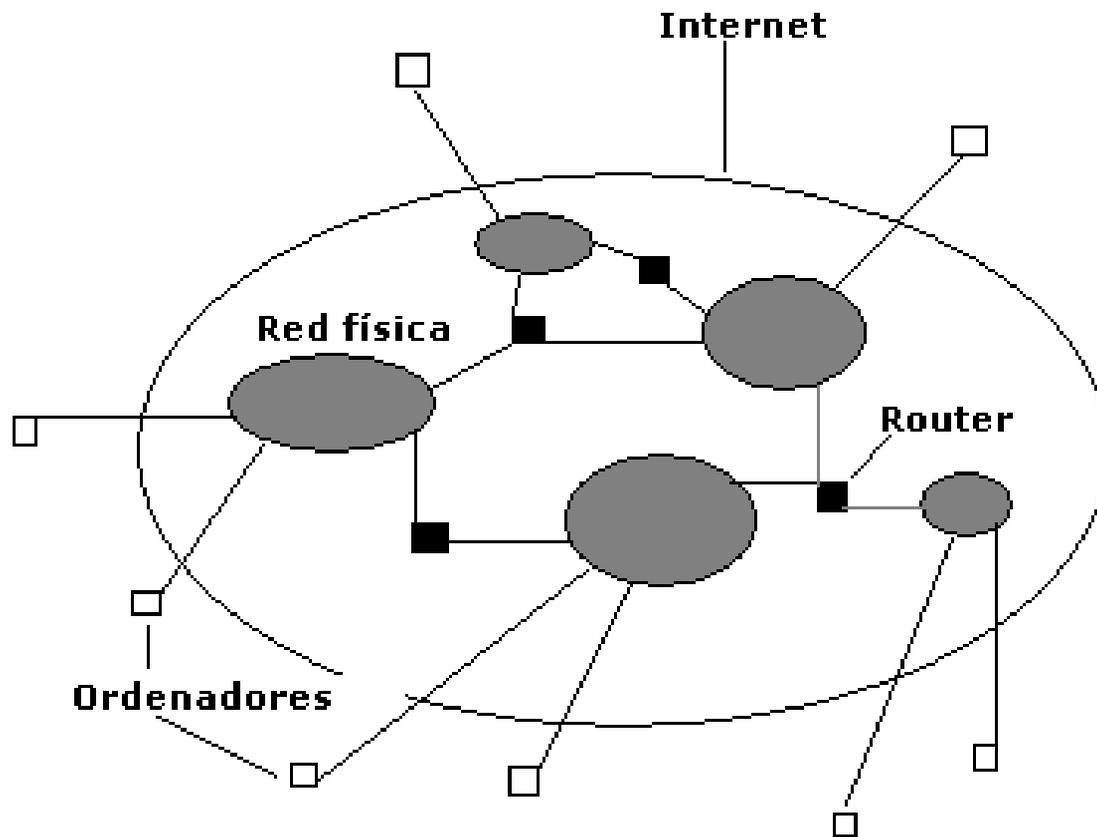
- **Capas:**

Capa Conceptual	Paso de objeto entre capas
1. Aplicación	Mensajes o flujos
2. Transporte	Paquetes de transporte
3. Red	Datagramas IP
4. Interfaz de red	Trama de red



- **Servicios a nivel de red**
 - Entrega de paquetes sin conexión
 - Transporte de flujo confiable
- **Servicios a nivel de aplicación**
- **Características de los servicios TCP/IP**
 - Independientes de la tecnología de red
 - Suministran interconexión universal
 - Proporcionan *acuses de recibo*

Conceptos básicos...: Arquitectura, enlace y tipos de redes



LAN

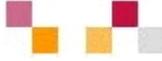
Cada ordenador tiene un dispositivo que le conecta al medio de red

WAN

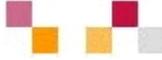
Red de commutadores de paquetes

Internet

Conjunto de redes conectadas por routers

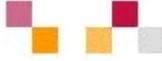


- **El paquete o datagrama es la unidad de información transmitida por TCP/IP.**
- **La información se trocea en paquetes para cumplir dos objetivos:**
 - **Compartir recursos.**
 - **Detectar y corregir errores.**
- **Cada hardware de red define un mecanismo de direccionamiento para especificar el destino de los paquetes.**
- **Todo dispositivo “conectado” tiene una dirección única: la que figura en los paquetes en el campo Dirección Destino.**
- **La identificación tiene dos partes: la de red y la de nodo**
- **Existe un doble direccionamiento:**
 - **físico (el de la tarjeta de red)**
 - **lógico (el de internet)**



- IP define el mecanismo de entrega sin conexión y no confiable.
- **Funciones:**
 - define el formato de datos: datagrama
 - selecciona la ruta de envío
 - concreta cómo nodos y routers deben procesar paquetes, generar mensajes de error y descartar paquetes
- **Formato del datagrama:**
 - Consta de 2 partes, encabezado y área de datos.
 - Los datagramas se transportan encapsulados dentro de las tramas de red. Si la trama es menor que el datagrama, se fragmenta en partes para transportarlo.

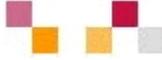
Versión	Longitud encabezado	Tipo de Servicio	Longitud datagrama	Identificación (Nº datagrama)	Desplazamiento fragmento	Flags	Tiempo de vida	Protocolo	Suma verificación	Dirección IP origen	Dirección IP destino	Opciones	Relleno	DATOS
---------	---------------------	------------------	--------------------	-------------------------------	--------------------------	-------	----------------	-----------	-------------------	---------------------	----------------------	----------	---------	-------



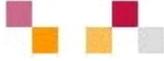
- **Direccionar es seleccionar un camino por donde mandar los paquetes. Lo realizan los "routers".**
- **Posibilidades en el envío:**
 - a) Destino en la misma red que el origen.....entrega directa
 - b) Destino en diferente red que el origen.....entrega indirecta
- **Si "b", el esquema de entrega es:**
 - trama – router – desencapsulamiento y lectura IP destino – encapsulamiento – envío al segundo router....., hasta la entrega directa
- **Para determinar la ruta, el router utiliza tablas de direccionamiento**

Red destino	Routers
147.96.x.x	entrega directa
otras	120.x.x.x

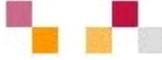
- **Consecuencias:**
 - Todo el tráfico para una red toma el mismo camino
 - Sólo el último router sabe si el nodo destino existe o está operando
 - Las rutas entre el nodo A-B no son bidireccionales a las de B-A



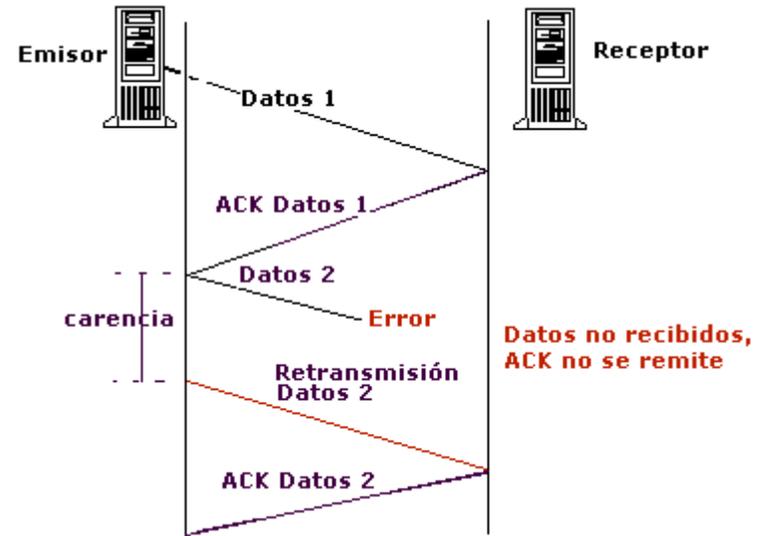
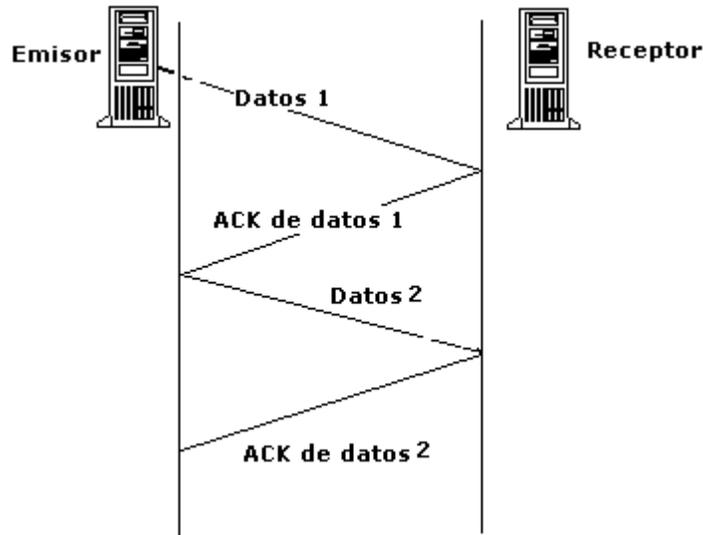
- **Cuando hay fallos en la comunicación, los routers envían mensajes de error o control a otros routers o nodos.**
 - fallos en líneas de comunicación, nodo destino desconectado, congestión de routers, etc.
- **El protocolo ICMP es el encargado de hacerlo**



- **¿Por qué es servicio?**
 - El intercambio de datos entre aplicaciones es muy grande.
 - Utilizar entrega sin conexión ni fiabilidad sería tedioso y requeriría que los programas incorporaran mecanismos de detección de errores. Es necesario, a nivel de protocolo de red, una entrega de flujo fiable que utilizan todas las aplicaciones.
 - TCP específica: el formato de datos y acuses de recibo, cómo distinguir el receptor en destino, forma de resolver errores, inicio de la transferencia, etc.
- **Funciones de TCP**
 - Transferencia básica de datos: el receptor recibe los datos en el mismo orden de envío.
 - Conexión virtual: identificada por los puntos extremos (nodo, puerto)
 - Transferencia con memoria intermedia
 - Flujo no estructurado
 - Conexión en ambos sentidos y control de flujo
 - Entrega fiable, mediante acuses de recibo ACK



- **Fases de la conexión TCP**



- **Mecanismos del control de flujo**

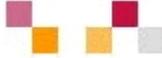
- El receptor informa al emisor del tamaño de su ventana
- Modificar temporizadores de retransmisión de segmentos
- Mecanismo PUSH para casos urgentes

Introducción a HTTP

- Es el protocolo de transferencia de hipertexto
- Se implementa en dos programas: cliente y servidor, que intercambian mensajes http
- Conceptos web:
 - Página web
 - Navegador o browser
 - Servidor web
 - Recursos
- Define cómo los navegadores piden recursos web a los servidores y cómo éstos los transfieren a los clientes.
- Esquema:
 - Vínculo - Mensaje *http* de petición del recurso - Mensaje http de respuesta con el recurso pedido.

Conexiones HTTP

- **No Persistentes (v 1.0)**
 - El servidor cierra la conexión en cuanto se suministra 1 recurso.
 - ejemplo 1 página con 10 imágenes = 11 conexiones y 22 ciclos RTT
- **Persistentes (v 1.1)**
 - El servidor deja abierta la conexión TCP tras enviar el primer recurso. Sucesivas peticiones emplean la misma conexión. Se cierra tras un intervalo sin peticiones.
 - Ejemplo: 1 pág con 10 imágenes en 1 conexión



- GET /directorio/pagina.html HTTP/1.1
- Connection: close
- User-agent: Mozilla/4.0
- Accept: text/html, image/gif, image/jpeg
- Accept-language:fr

Método	SP	URL	sp	Versión	CR	LF	Línea de petición	
cabecera 1	:	valor 1	CR	LF			Líneas de cabecera	
.....								
cabecera N	:	valor N	CR	LF				
CR	LF							
Cuerpo del mensaje								



- HTTP/1.1 200 OK
- Connection: close
- Date: Thu, 06 Aug 1998 12:00:15 GMT
- Server: Apache/1.3.0 (Unix)
- Last-Modified: Mon, 22 Jun 1998 09:23:24 GMT
- Content-Length: 6821
- Content-Type: text/html
- data data data data data ...

Versión	SP	Código Status	SP	Frase	CR	LF	Línea de estado
cabecera 1	:	valor 1	CR	LF			Líneas de cabecera
.....							
cabecera N	:	valor N	CR	LF			
CR	LF						
Cuerpo del mensaje							

- Principales códigos:
 - 200 OK
 - 301 Traslado permanente
 - 400 Petición errónea
 - 404 No encontrado
 - 500 Error del servidor
 - 505 Versión http no soportada

Interacciones con el usuario: autenticación y cookies



- **Autenticación: solicitar al usuario sus credenciales de acceso.**
- **Cookies. Mecanismo para seguir la pista de un usuario**



- **Objetivo de la OAI**
 - Transformar el modelo de comunicación académica
 - ¿Cómo? Definiendo principios organizativos y técnicos de un modelo de publicación académica abierto.

- **Orígenes de la OAI**
 - Archivos electrónicos de preprints (e-print)
 - e-print: vehículos informales de difusión de la investigación

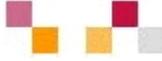
- **Corrientes de la OAI**
 - **La política: open=acceso público y gratuito a la producción académica**
 - . La investigación pertenece a la comunidad académica
 - . La gratuidad se alinea con los intereses de autores y lectores
 - . Los e-prints aceleran y democratizan la investigación
 - . Es la tendencia subrayada por la BOAI y la Declaración de Berlín

 - **La tecnocrática: open=incremento de la interoperatividad entre e-prints**
 - . Interfaces que aumenten los proveedores de la información dispuestos a ofrecer su contenido



- **Principales acuerdos**
 - Norma de metadatos para recuperar documentos: Dublin Core
 - XML como sintaxis
 - DC para extraer metadatos
 - Sistema de identificación uniforme

- **Valoración de las decisiones**
 - Rechazo de la búsqueda distribuida
 - Traslado de la complejidad a los servicios



- Metadatos

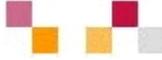
- Todos los proveedores ofrecen DC sin calificar (interoperatividad)
-
- OAI sustenta cualquier conjunto de metadatos si tiene sintaxis XML y un esquema de validación (extensibilidad)

- Funcionalidades

- 2 Participantes: proveedores de datos y de servicios
- Transacciones sobre *http* entre el harvester y el repositorio
- Recolección selectiva: por fecha y por subcolecciones
- Respuestas en *XML*
- No se contempla: autenticación, frecuencia recolección, uso aceptable de metadatos

- Registros OAI

- Header: identificador del registro y datestamp (fecha de creación...)
- Metadatos: descripciones de los registros en DC

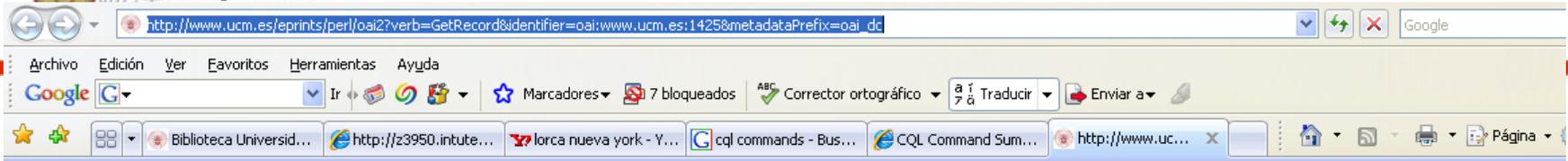


- **Estructura de una petición**

- **Url-base:** host y puerto del servidor e-print
- **Argumentos:** órdenes con formato clave = valor

- **Órdenes OAI**

- *GetRecord.* Recuperar un registro individual de un ítem
- *Identify.* Recupera información de un depósito: nombre, URL, versión, email del administrador.
- *ListIdentifier.* Lista todos los ids únicos que corresponden a los registros del archivo
- *ListMetadataFormats.* Formatos de metadatos que soporta el archivo
- *ListRecords.* Recupera metadatos para varios registros
- *ListSets.* Listado jerárquico de colecciones del archivo



```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd" xmlns="http://www.openarchives.org/OAI/2.0/">
  <responseDate>2007-09-14T08:57:38Z</responseDate>
  <request verb="GetRecord" identifier="oai:www.ucm.es:1425" metadataPrefix="oai_dc">http://www.ucm.es/eprints/eprints/perl/oai2</request>
- <GetRecord>
- <record>
  - <header>
    <identifier>oai:www.ucm.es:1425</identifier>
    <datestamp>2004-11-15</datestamp>
    <setSpec>747970653D626F6F6B5F73656374696F6E</setSpec>
  </header>
  - <metadata>
    - <oai_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:title>La biblioteca pública en la sociedad de la información</dc:title>
      <dc:creator>Magán Wals, José Antonio</dc:creator>
      <dc:subject>Bibliotecas</dc:subject>
      <dc:description>La biblioteca pública es un espacio imprescindible para el desarrollo humano y de los derechos humanos. A fin de poder cumplir con
        su función (preservación y difusión del conocimiento, la información y facilitar la expansión de los valores democráticos) los bibliotecarios han
        desarrollado nuevos servicios y se adaptan a las circunstancias actuales de la sociedad de la información. En este contexto es imprescindible la
        colaboración entre la biblioteca pública y las bibliotecas nacionales, universitarias y escolares a fin de crear verdaderos sistemas de información
        públicos.</dc:description>
      <dc:publisher>Javier Pérez Iglesias</dc:publisher>
      <dc:date>2004-11-10</dc:date>
      <dc:type>Sección de Libro</dc:type>
      <dc:type>NonPeerReviewed</dc:type>
      <dc:identifier>http://www.ucm.es/eprints/1425/</dc:identifier>
      <dc:format>application/octet-stream</dc:format>
      <dc:relation>http://www.ucm.es/eprints/1425/01/La%5Fbiblioteca%5Fp%C3%BAblica%5Fen%5Fla%5Fsociedad%5Fde%5Fla%
        5Finformaci%C3%B3n.doc</dc:relation>
      <dc:format>text/plain</dc:format>
      <dc:relation>http://www.ucm.es/eprints/1425/02/La%5Fbiblioteca%5Fp%C3%BAblica%5Fen%5Fla%5Fsociedad%5Fde%5Fla%
        5Finformaci%C3%B3n.txt</dc:relation>
    </oai_dc:dc>
  </metadata>
</record>
</GetRecord>
</OAI-PMH>

```



- **Objetivos:**
 - **Facilitar el uso y reutilización de datos**
 - **Normalizar la sintaxis de las consultas en las búsquedas por Internet**
 - **Normalizar la estructurar las respuestas .**
 - http://search.yahoo.com/search;_ylt=A0geu8yoO.pGj0EBVXBXNyoA?p=lorca+nueva+york&fr=fp-pull-web-t
 - <http://search.msn.com/results.aspx?q=lorca+nueva+york&form=QBRE>
- **SRU y SRW son protocolos basados en web services.**
 - **SRU utiliza sintaxis REST**
 - Transmisión vía http, consulta con cadenas nombre/valor y respuesta en XML.
 - **SRW utiliza sintaxis SOAP**
 - Transmisión por otros mecanismos de transporte: http, mail, telnet, etc.
 - Consulta y respuesta codificada en XML.
- **Operaciones SRU/W**
 - **Explain**
 - **Scan**
 - **SearchRetrieve. Utiliza el lenguaje de consulta CQL**
- **Ejemplos: COPAC, INTUTE**

[http://copac.ac.uk:3000/copac?query="Garcia%20Lorca"&operation=searchRetrieve&version=1.1&maximumRecords=3](http://copac.ac.uk:3000/copac?query=)

Archivo Edición Ver Favoritos Herramientas Ayuda

Google Ir

Marcadores 7 bloqueados Corrector ortográfico Traducir

Biblioteca Universid... <http://copac.ac...> Intute - SRU/SRW lorca nueva york - Y... cql commands - Bus...

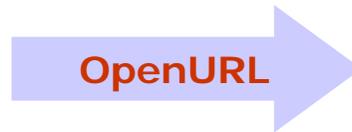
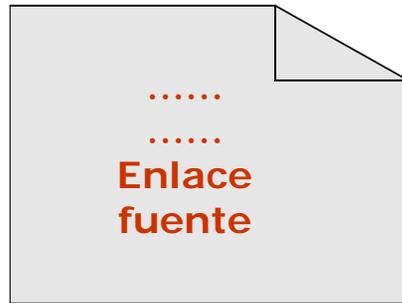
```

<?xml version="1.0" ?>
- <zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>2558</zs:numberOfRecords>
- <zs:records>
- <zs:record>
  <zs:recordPacking>xml</zs:recordPacking>
- <zs:recordData>
- <mods xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema"
  xmlns:h="http://copac.ac.uk/schemas/holdings/v1" xmlns:cpc="http://copac.ac.uk/schemas/mods-copac/v1"
  xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-copac/v1
  http://copac.ac.uk/schemas/mods-copac/v1/mods-copac-v1.xsd http://copac.ac.uk/schemas/holdings/v1/holdings-v1.xsd">
- <recordInfo>
  <recordIdentifier source="copac">72004174995</recordIdentifier>
  <recordContentSource>Copac</recordContentSource>
  <recordCreationDate>20060525</recordCreationDate>
</recordInfo>
- <originInfo>
  <dateIssued encoding="marc">1972</dateIssued>
</originInfo>
- <originInfo>
- <place>
  <placeTerm type="text">Ancona, Milano :</placeTerm>
</place>
  <publisher>Edizioni musicali Bèrben,</publisher>
  <dateIssued>[1972]</dateIssued>
</originInfo>
- <titleInfo>
  <title>Seis canciones de Federico García Lorca. Op. 8. For voice and guitar, etc.</title>
</titleInfo>
- <name type="personal">
  <namePart>Apivor, Denis,</namePart>
  <namePart type="date">1916-</namePart>
- <role>
  <roleTerm type="text">creator</roleTerm>

```



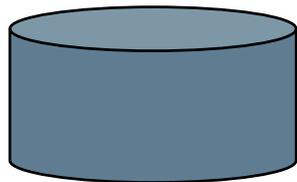
Recurso de Información



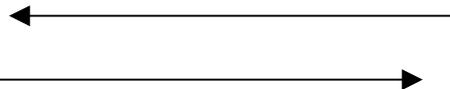
Servidor de resolución



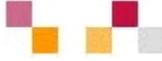
- . Directamente, del OpenURL
- . Indirectamente, del identificador



Base de datos de metadatos



Servicios complementarios de enlace (*extended services*)



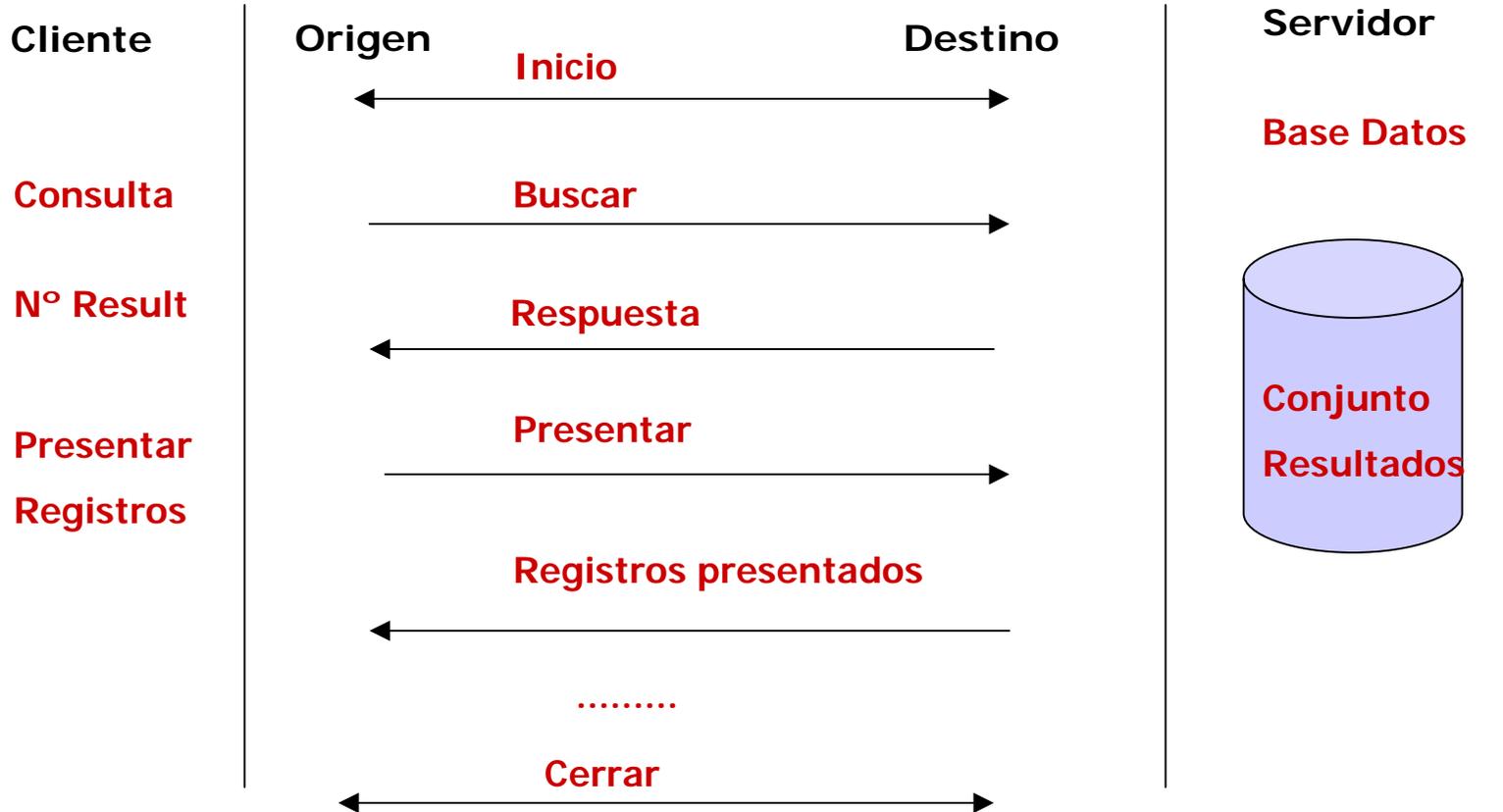
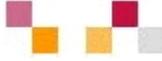


- **Un protocolo que permite buscar información entre diferentes sistemas y utilizar un único interfaz de usuario.**
- **Los usuarios de un sistema puede buscar información en otro sistema sin necesidad de conocer cómo funciona el sistema remoto.**



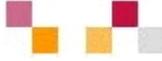
- **Un usuario de un Opac selecciona un servidor Z destino (normalmente un catálogo de otra biblioteca) desde un menú del Opac.**
- **Introduce un término de búsqueda. El soft del Opac envía los términos de búsqueda y los detalles del destino a un cliente Z39.50 que corre como un módulo integrado del Sistema Bibliotecario.**
- **El cliente Z traduce los términos de búsqueda al lenguaje Z y contacta con el soft del servidor Z destino, realizan una negociación preliminar y establecen una Asociación-Z entre los dos sistemas.**
- **El servidor Z traduce la consulta en lenguaje Z a su base de datos y recibe la respuesta con el número de registros pertinentes**
- **El cliente Z recibe los registros (tras pedirlos) y los presenta en el interfaz del opac del usuario.**



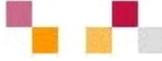




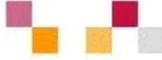
- **Acceder a recursos multidisciplinares: catálogos de bibliotecas, museos, información estatal, datos geoespaciales, bases de datos.**
- **Proporcionar servicios:**
 - Creación de catálogos colectivos virtuales
 - Captación de registros bibliográficos;
 - Préstamo interbibliotecario
 - Servicios de DSI y creación de perfiles (v.3)
- **Ocultar la tecnología a los usuarios, que sólo tienen que preocuparse de buscar contra una gran base y saber utilizar su propio cliente Z.**
- **Ofrece un interfaz uniforme a diferentes sistemas**
- **Una sólo búsqueda contra múltiples recursos: consultas multidifusión (a más de un servidor).**



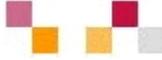
- **Clientes integrados en el Sistema de Gestión de la Biblioteca**
- **Clientes autónomos comerciales (End Note, Bookwhere, Znavigator...)**



- **Raíces en los 70s: iniciativa LSP (Linked Systems Project, Proyecto de Sistemas Enlazados).**
- **1988. NISO adopta como norma el protocolo empleado por la LSP**
- **1989. NISO nombra a la LC como Agencia Oficial de Mantenimiento y Registro de la norma.**
- **1992. Versión 2 de la norma**
- **1995. Versión 3. Adoptada por la ISO como ISO23950 en 1998**



- Inicio
 - Búsqueda
 - Recuperación
 - Borrar
 - Control de acceso
 - Control del recurso y contable
- Básicos**
- Ordenar
 - Hojear
 - Servicios complementarios
 - Explicación



- **Forma normalizada de implementar opciones e interpretar el estándar. Proporcionando así resultados más consistentes.**
- **Significan una fragmentación de la comunidad Z en subcomunidades aisladas.**
- **Responde a los problemas de interoperatividad derivados de las características opcionales o incompletamente especificadas por la norma.**
- **Respuesta a la carencia de procesos correctamente articulados para establecer normas para conjuntos de atributos y sintaxis de registros que soporten diferente semántica para los distintos RI.**
- **Un perfil es una especificación concreta de las características y funciones del protocolo Z a la que es conforme una implementación, es decir, es una aplicación determinada del protocolo restringiendo su versatilidad para mejorar el éxito en la recuperación de inf.**