

The author is not a native speaker of Spanish. I apologize for any misspellings, typos, incorrect grammar or just plain wrong use of the Spanish language. I do recognize that there are many words throughout this document that do not have the proper diacritical marks (í, é, ó, ñ, etc.) This is a function of Microsoft Word's inability to deal with Unicode and the difficulty of adding each mark in, one by one. If anyone wants to send me corrections of my Spanish, it would be greatly appreciated and taken as a kindness and professional courtesy. –Carol Hixson

## Slide 1 **Técnicas**

Buenos días. Estoy muy contenta de estar aquí con todos vosotros. Os tengo que confesar que esta es la primera vez que visito Barcelona. Llegué el lunes y estoy aprovechando cada momento libre que tengo para visitar museos, monumentos y conocer un poquito vuestra ciudad que desde ya principio me ha parecido hermosísima. Tenéis una gran suerte de vivir en una ciudad tan amistosa. También tengo que agradecerlos por haberme invitado a hablar con vosotros sobre los repositorios institucionales y nuestra experiencia a la Universidad de Oregon. Si teneis dudas o alguna inquietud durante o posterior a mi exposición por favor no dudéis en pedirme la aclaracion respectiva, ¿vale?

## Slide 2

Esta sesión cubrirá algunos aspectos técnicos de como establecer y mantener un repositorio institucional usando como ejemplo el nuestro, el Scholars' Bank. Voy a hablar un poquito de la plataforma de programario, las modificaciones locales del programario, procesos de *backup*, registro y preservación a largo plazo. Media hora no es mucho por hablar de todos los aspectos

técnicos. En las otras dos sesiones también hablaré un poco de otros aspectos técnicos que les ayudara a formarse una idea más acabada del tema.

### **Slide 3**

Normalmente se concluye con los agradecimientos. Sin embargo, yo he preferido comenzar con estos dado que cuento con un excelente equipo colaborador. En primer lugar, agradezco la ayuda brindada por uno de los bibliotecarios de mi departamento, Corey Harper. Corey tiene un gran conocimiento técnico y fue él quien me dió muchos de los detalles técnicos a los que me referire durante mi presentación. Yo utilizo el programario y entiendo bastante bien los retos y limitaciones de este. Sin embargo, es mi colega Corey la persona que hace funcionar el servidor y quien actualiza el programario cuando hay nuevas instalaciones. Corey se sumó al equipo en la primavera del año pasado y es él quien entiende porque funciona o no el servidor.

### **Slide 4**

También quiero agradecer a Álvaro Mardones Llanos, chileno y estudiante graduado en Estudios Internacionales a la Universidad de Oregon quien me ayudó con mi español. Vuestra colega Alice Keefer también me ayudó con alguna terminología. Pido vuestra paciencia con esta sesión a causa de las dificultades con la jerga técnica (no sé si hemos identificado la palabra exacta algunas veces). Y en general, tengo que pedir vuestra paciencia con mi castellano.

Yo estudié castellano en la universidad hace casi 30 años. Yo nunca he vivido en un país de habla española. He utilizado el español en mi trabajo cuando yo era una catalogadora de materias escritas en español. Hoy en día no tengo mucha oportunidad ni de leer castellano ni de hablarlo. Con la ayuda brindada por Álvaro espero poder explicaros medianamente claro lo que hacemos en la

Universidad así como lo que pensamos de nuestro esfuerzo y del esfuerzo mundial del que vosotros también sois una parte.

También, quiero explicaros que después de estudiar español, estudié portugués y viví en el Brasil. Muchas veces pienso con las palabras portuguesas, o también francesas, que es otro idioma que estudio bastante puesto que tengo familia ahora en Francia. Espero que podáis entender lo que quiero decir. Si no, pues os animo a preguntarme pues mi idea es tratar el tema lo mejor posible objeto no existan dudas.

## **Slide 5 La Universidad de Oregon**

La Universidad de Oregon es la mayor Universidad en el estado y la única del primer nivel. La Universidad de Oregon cuenta con más de dieciocho mil (18,861) estudiantes y más de ciento veinte (129) programas de estudio.

## **Slide 6 Las Bibliotecas de la Universidad**

Hay siete bibliotecas en el sistema de la Universidad de Oregon y es el sistema mayor en el estado. Tenemos más de 2.5 millones de volúmenes catalogados y más de 500,000 otros recursos todavía no catalogados, por ejemplo muchas cartas, los archivos de la universidad, colecciones de fotografías y manuscritos.

Trabajan en las siete bibliotecas más de 50 bibliotecarios, casi 100 otros empleados, y más de 60 estudiantes

## **Slide 7 Servicio de Biblioteca Digital y Metadatos**

Mi departamento tiene 30 empleados, incluyendo seis bibliotecarios. Nosotros tenemos la responsabilidad por todo tipo de colección digital, incluyendo el repositorio institucional, y también la catalogación y la preservación de los recursos tradicionales (libros, periódicos), y el servicio de microfilm.

## **Slide 8 Los comienzos**

Al comienzo, nuestro equipo consistía en mi persona y 4 otros bibliotecarios encargados de investigar las posibilidades de establecer un repositorio institucional. Es el caso que sólo uno de nosotros tenía gran aptitud técnica. (Y al principio este no era Corey sino que otro bibliotecario de otro departamento) Yo tenía alguna aptitud técnica pero los otros miembros del grupo no tenían ninguna. !Immaginese lo que significa trabajar en esas condiciones!

Existía la esperanza de que nuestro equipo sería capaz de investigar e implementar sin la ayuda de otros un sistema inicial para indagar las posibilidades de un repositorio institucional.

Después de más de dos años de esfuerzo, nuestro equipo hoy está compuesto por siete bibliotecarios. Por mi persona, el director del centro de educación técnica académica de la biblioteca, la directora del departamento de referencia e instrucción, una de las bibliotecarias nuevas en su departamento, la archivista de la universidad (quien es también abogado), mi colega Corey que tiene bastante aptitud técnica, y ahora otra bibliotecaria de mi departamento para ayudarme con el contacto con la facultad y estudiantes y con la administración del sitio.

## **Slide 9    Requisitos del programario**

Nuestros requisitos básicos del programario eran:

- Permitir la fácil creación, uso y administración de los documentos digitales distribuidos en Internet
- (con una Interfaz de usuario basada en la web)
- Facilitar la creación de colecciones de documentos en distintas áreas
- Almacenar cualquier tipo de documentos y realizar búsquedas en base a metadatos
- Ser construido utilizando componentes y tecnologías estándares y abiertos
- El producto final debe ser de dominio público y abierto

## **Slide 10    Otros requisitos del programario**

Puesto que nuestro equipo era pequeño y puesto que todos nosotros teníamos muchas otras responsabilidades, buscábamos un programario que reuniera estos criterios:

- Que otras instituciones lo hubieran probado para que nosotros no tuvieramos que ser los primeros en utilizarlo
- Que fuera fácil de implementar (puesto que teníamos que hacerlo sin la asistencia de terceros)
- Que se conformara con los estándares del Open Archives Initiative-Protocolo para el Cosechamiento de Metadatos (OAI-PMH) para poder ajuntarnos con los otros repositorios,
- Que no necesitara gran inversión en equipo (hardware) – teníamos acceso a un servidor viejo utilizado antiguamente para otro programario
- Y finalmente, que el programario fuera barato puesto que no teníamos mucho dinero – que es siempre nuestra realidad en el estado de Oregon

## **Slide 11 Investigación del programario**

Nosotros escogimos DSpace porque satisfacía todos estos criterios.

Como vosotros sabéis, Dspace es un programario que facilita la creación, uso y administración de repositorios de documentos digitales en Internet

## **Slide 12 Requisitos, retos, ventajas**

Voy a discutir los Aspectos Técnicos del programario en dos partes:

- Primero, voy a relacionar algunos detalles sobre la instalación y el funcionamiento del programario. Eso son los detalles que me ha dado mi colega Corey.
- Voy a explicar también nuestro ambiente técnico.
- Después, propongo hablar de nuestro uso del programario para continuar con la instalación así como los problemas y soluciones que nosotros hemos descubierto.

## **Slide 13 Hardware (servidor)**

El hardware que nosotros utilizamos consiste en:

- Procesador con 1.30 Ghz
- 512 MB de memoria RAM
- 40 GB de disco duro

## **Slide 14 Programarios**

Para nuestro repositorio utilizamos:

- Dspace 1.2.2
  - Intentaremos instalar 1.3 en octubre

Para hacer funcionar Dspace, es necesario además utilizar otros programarios. Nosotros utilizamos los que se ve aquí en la diapositiva.

## **Slide 15 Base de datos**

La base de datos puede ser Oracle o Postgresql. Nosotros utilizamos Postgresql. Debe tenerse en cuenta que DSpace no funcionaría con MySql.

Como dice la documentación de DSpace, se puede ofrecer la mayoría del funcionamiento con cualquier base de datos SQL estándar o que apoye las transacciones. Ahora, los índices de “browse” utilizan algunos caracteres específicos a PostgreSQL. Si se quiere utilizar otra base de datos, se tendría que modificar el código de DSpace a fin de hacerlo funcionar completamente.

## **Slide 16 Requisitos técnicos para instalación**

Segundo Corey, los requisitos técnicos para instalar DSpace son los que se ven aquí (no voy a leerlos)

## **Slide 17 Ventajas técnicas de DSpace**

También Segundo Corey las ventajas técnicas DSpace son que:

- El Código es limpio (o puro) y Utiliza los estándares
- Además del código basado en los estándares, la arquitectura también se base en estándares como Dublin Core, METS, OAI, CNI Handles para la persistencia de los URL, y otros.
- Código se encuentra bien comentado. Se puede ver lo que el programador ha hecho y se puede modificar o corregir fácilmente.
- Buena arquitectura –estrato de negocios (business layer) separado del estrato de la aplicación.
- Postgresql es poderoso y robusto
  - Es el mejor sistema de base de datos de libre distribución.
  - Se puede pesquisar la base de datos directamente de la línea de mando para extraer información. (que hacemos muchas veces)
- La base de datos permite las búsquedas normales de SQL y hace la corrección de problemas más fácil

## **Slide 18 Retos técnicos**

No hay ninguna manera fácil de lograr que Tomcat pueda listar los puertos de norma. Nosotros utilizamos modjk2 para corregirlo. Apache ha decidido terminar de sostener modjk2. Ello no sería un problema para nosotros si no se modifican mucho ni Tomcat ni Apache. La comunidad de Dspace está hablando de dejar de utilizar Java. Si lo hacen, no necesitaríamos Tomcat.

## **Slide 19 : Irtest**

Nosotros tenemos una Instalación de prueba en la que experimentamos. Cuando hay nueva versión, la instalamos primero aquí.

## **Slide 20 Exportación de los metadatos OAI-PMH**

Antes señalé que un requisito nuestro fue que el repositorio sosteniese los estándares de OAI-PMH. Como ya sabéis, la OAI establece la creación de repositorios abiertos y distribuidos que contengan al menos los metadatos de los recursos que albergan o describen. Nosotros aseguramos que podíamos exportar los metadatos conforme con los requisitos de OAI. Aquí podéis observar una pantalla del resultado. Como podéis apreciar, la pantalla de resultado nos ha presentando todas las identificaciones de las comunidades y colecciones de Scholars' Bank.

## **Slide 21 Preservación**

La preservación constituye el problema más serio que tenemos en la actualidad. ¿Vamos a preservar sólo los dígitos binarios? O bien ¿vamos a preservar el contenido? Tuvimos una discusión muy larga en nuestro equipo sobre este particular. La decisión que tomamos finalmente es que intentaremos preservar los dos – los dígitos y el contenido. Pero no vamos a comprometernos a ello. Hay varias gestiones que seguimos para tratar de proteger la integridad y la utilidad de los ficheros y en esta página las explicamos.

## **Slide 22 registro de metadatos**

MIT y Hewlett Packard, los que han desarrollada DSpace, dicen que DSpace sostiene la preservación. El registro de metadatos que existe en el programario facilita la preservación, pero es una parte muy pequeña de este. También es importante tener políticas que determinan el contenido de los campos de dato.

## **Slide 23 registro de los formatos de bitstream**

También el registro de los formatos de bitstream constituye una parte de un programa de preservación. Sin embargo, se necesita establecer políticas para tener un programa sólido de preservación. Nosotros tenemos un grupo llamado los Coordinadores del Contenido Digital que establece políticas sobre la preservación de todo tipo de contenido digital que creemos o obtenemos en la biblioteca.

## **Slide 24 verificación de los checksums**

Otra parte de un programa de preservación lo constituye la verificación de los checksums. **(click)** Nosotros todavía no tenemos un programa automático para comprobar los checksums – esto es, para verificar que el fichero efectivamente no tiene defectos. Hemos hablado de la necesidad de comenzar un programa. Ahora lo hacemos de vez en cuando, pero esto es completamente insuficiente.

## Slide 25 respaldo de cinta

Una parte muy importante es el backup – el respaldo. Nuestro respaldo utiliza la cinta. Nuestro sistema es el de respaldar el base de datos en cinta con esta frecuencia que se ve aquí: **(don't read)**

- Hace una semana (one week ago)
- Hace dos semanas (two weeks ago)
- Hace tres meses (three months ago)
- El diez de julio del previo año fiscal

## Slide 26 vaciado de datos

También hacemos vaciado de datos que residen en el espacio de respaldo en el servidor mismo,

Cada cinta de respaldo contempla también los vaciados de datos, tan bien como la copia dinámica de la memoria del servidor completo en el día del respaldo. Esto incluye también los ficheros.

No se puede normalizar o restaurar Dspace de la cinta. Por eso también hacemos los vaciados de datos. El horario del vaciado de datos es lo que se ve aquí **(don't read)**

- Cada día
- Los últimos siete días
- El primer día de cada mes
- Los tenemos desde el noviembre de 2004

## Slide 27 modificaciones locales

Hemos hecho muy pocas modificaciones locales del programario.

Cuando se utiliza un programario libre de “open source” (fuente abierta), siempre se nos impone el deber de decidir si vamos a aceptar el diseño programado o no. En la biblioteca decidimos utilizar el diseño programado, esto es, tal cual éste era originalmente. Nosotros hicimos muy pocas modificaciones locales al diseño. La decisión se basó en los siguientes puntos:

- primero, todos estábamos de acuerdo con la visión principal;
- segundo, no teníamos personas ni con la experiencia necesaria ni con suficiente tiempo para modificar el diseño básico. Ya he explicado la composición de nuestro equipo.
- Finalmente, en ese entonces todavía no sabíamos si íbamos a seguir utilizando DSpace o si íbamos aún a continuar con el repositorio.

Hemos hecho algunas modificaciones pequeñas en las áreas de estadísticas, remisión y búsquedas

Sólo en las estadísticas y búsquedas hemos modificado un poquito el programario. La modificación acerca de la remisión es sólo un “workaround” donde contravinimos el diseño un poco.

## **Slide 28 estadísticas**

Para manejar y promover nuestro repositorio, queremos saber una variedad de cosas, como: quién visita el repositorio, cuántas veces se visita, cuáles ficheros se utilizan, como han descubierto el repositorio, de dónde son los usuarios, quién ha remitido los ficheros, qué formatos se representan en el repositorio, cuándo eran remitidos los ficheros, y más.

DSpace no tiene herramientas administrativas para proporcionarnos la mayoría de la información que queremos saber. Nosotros utilizamos un programario llamado Webalizer para recoger y analizar varias estadísticas, como por ejemplo las estadísticas sobre el uso del repositorio. Webalizer analiza los registros estándares de un servidor. Nos da la información básica sobre quién visita nuestro sitio de Web, que tipo de navegador utilizan, de dónde son referidos, cuáles páginas ellos ven, etc.

Con demandas a la base de datos, usando mysql, nosotros tenemos información para poder interpretar las estadísticas. DSpace 1.3, que vamos a instalar probablemente en octubre, tiene algunas herramientas estadísticas. Es probable que sigamos utilizando una variedad de herramientas para descubrir y analizar las estadísticas.

## **Slide 29 consultas al servidor**

Este es un resultado de Webalizer, mostrando un resumen de las estadísticas – Aquí se pueden ver los golpes o consultas al servidor.

## **Slide 30 estadísticas diarias**

Webalizer también nos da las estadísticas de manera más detallada, como el uso diario.

## **Slide 31 países de origen**

Webalizer también nos da información sobre los países de origen de nuestros usuarios. Hasta hoy, hemos recibido visitas de más de 70 países. Webalizer nos proporciona la misma información en forma gráfica.

## **Slide 32 Estadísticas de Dspace 1.3**

Las Estadísticas de Dspace 1.3 aparecen de esta manera si no se modifica el código, con sólo el handle (**arrow**) y el número de consultas (**2<sup>nd</sup> arrow**)

El nuevo código de DSpace tiene un fichero de configuración (config) para el módulo de las estadísticas. (/dspace/conf/dstat.cfg) Este fichero tiene una variable, item.lookup, que tiene asignación implícita de cinco. Nosotros lo cambiamos a quinientos. (**Next slide**)

## **Slide 33 Estadísticas de Dspace 1.3**

Hacemos eso para que se pueda apreciar no sólo el handle sino también el título que ha sido visto. (**arrow**) Nuestra preocupación con este cambio es que la demanda se basa en java y esto podría presentar un problema de ejecución si las estadísticas se compilan cada noche o cada semana. Pensamos hacer públicas estas estadísticas pero vamos a vigilar su uso cuidadosamente.

## **Slide 34 pantalla de remisión**

Dspace permite la modificación de la pantalla de remisión. Es bastante simple modificarla. Se puede proveer información que deber aparecer cada vez que se remite un documento a una colección. Por ejemplo, aquí hemos provecho un tema (**1<sup>st</sup> arrow**), un título (**2<sup>nd</sup> arrow**), y el “tipo” (entrega de un periódico) (**3rd arrow**)

## **Slide 35 problemas con la pantalla de remisión**

Pero, este “tipo” no aparece en la pantalla (**click**) de remisión porque no es un tipo definido por el programario. Para aceptar un nuevo término para el tipo de remisión, hay que modificar el código. Hemos decido no modificar el programario.

Este es sólo un ejemplo de los problemas con la forma de remisión.

## **Slide 36 Remisión facilitada**

Puesto que más de noventa y cinco (95) por ciento del contenido actual fue enviado o remitido por empleados de la biblioteca con el permiso de los autores, hemos desarrollado algunos métodos para contravenir el proceso normal de remitir ficheros en DSpace.

Adquirimos un email o un permiso en papel del autor y lo convertimos en un fichero de texto simple (con la extension de .txt). Lo ajuntamos a la remisión, junto con el fichero de permiso automático de DSpace que registra la dirección de email de quien lo ha remitido. Aquí se ve los dos ficheros de permiso.

Cuesta bastante trabajo hacer ésto pero lo consideramos necesario para registrar el permiso junto con la remisión.

## **Slide 37 problema con las búsquedas**

Uno de nuestros usuarios descubrió un problema con las búsquedas dentro de una colección. Cada vez que él obtuvo un grupo de resultados de más de una pantalla – como Aquí (**flecha**) – marcando el botón para la próxima página, resultó en una búsqueda de todo el archivo, en vez de continuar con los resultados específicos para la colección, como sería lo más lógico.

El problema estaba en el fichero results.jsp. Nuestra reacción fue bastante simple. Corey agregó una variable, llamada searchScope, estableció la variable de acuerdo a otras variables en una demanda de http y usó esa variable para modificar la creación de enlaces para los próximos, previos, y numerados botones de páginas.

Contribuimos esta modificación del código a la comunidad de DSpace y ahora es parte de DSpace 1.2.2

## **Slide 38 Problemas con los metadatos**

Uno de los problemas tiene relación con los metadatos y las limitaciones de Dublin Core, especialmente dentro de DSpace. La documentación de DSpace señala que apoya Dublin Core simple y modificado (qualified). Y dentro de la interfaz administrativa se puede elegir los elementos modificados. Por ejemplo,

## **Slide 39 Modificación de los metadatos**

El modelo normalizado de Dspace es el de mapear tema a DC Subject – Tema simple sin modificación, como se puede apreciar aquí. (**click**)

## **Slide 40 Modificación de los metadatos**

Aquí se pueden apreciar los temas utilizados para una remisión en el archivo. Son temas del sistema de LCSH – la Biblioteca del Congreso de los EEUU. (**arrow**)

## **Slide 41 Modificación de los metadatos**

Se puede modificar la correlación o conducción (mapping) del elemento de Dublin Core-- en este caso para representar el sistema de temas, como los temas de la biblioteca de congreso de los EEUU. (**click**) Como se puede ver aquí.

## **Slide 42 Modificación de los metadatos**

Pero después de modificar la correlación (mapping) de los metadatos de DC Subject a DC Subject.lcsh  
Nos encontramos con la situación en que no se puede ver ni buscar este tema. Simplemente desaparece (**flecha**)

Este es un ejemplo en que, en vez de modificar el código, nosotros aceptamos las limitaciones del programario.

Hasta ahora no ha sido un problema tan grande para nosotros.

## **Slide 43 problemas con la remisión**

Muchas veces nos encontramos con problemas con la remisión – el programario no las acepta. Todavía no tenemos ninguna explicación ni razón del porque se presenta esta situación. Sin embargo, cuando lo intentamos otra vez, el proceso funcionó.

Es un misterio.

## **Slide 44 problemas con el registro del permiso**

También tenemos bastante problemas con el registro del permiso. Como con la remisión tampoco sabemos porque esto tiene lugar. Generalmente la próxima vez que lo hacemos, funciona. El programario todavía tiene algunos defectos, pero no son bastante serios para que nosotros nos demos el tiempo a modificar el código.

## **Slide 45 Información de contacto**

Esto es todo que quiero decir sobre los aspectos técnicos. Tenemos algunos minutos para una discusión. Si vuestras preguntas son muy técnicas yo las traspasaré a mi colega Corey después de volver.