

Indicadores de calidad en repositorios de acceso abierto

Dra. Marisa R. De Giusti

PREBI-SEDICI

Universidad Nacional de La Plata

Semana de Acceso Abierto

Del 23 al 29 de octubre de 2017

Argentina - Jueves 26 de octubre



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](#)

Objetivos de esta charla



Reflexionar sobre qué se entiende por calidad en los repositorios de acceso abierto: ¿dimensiones, aspectos, parámetros?

La definición de la Norma ISO 9000 no deja dudas: “**Calidad:** *grado en el que un conjunto de características inherentes a un objeto (producto, servicio, proceso, persona, organización, sistema o recurso) cumple con los requisitos.*”

El requisito es un problema porque es la expectativa del usuario, del cliente.

Ahora ¿de qué modo “se mide” esa calidad?



Calidad en el mundo de los repositorios de AA

Después de 15 años de declaraciones sobre el acceso abierto, es bueno pensar cómo los repositorios sirven a los usuarios, investigadores, docentes, ¿qué se percibe?, ¿qué servicios se brindan?.

Pensar de algún modo parámetros e indicadores cualitativos y de los otros para objetivar esa percepción de calidad y ver si se cumple con las premisas del AA.



Requisitos

Entendidos como necesidades o expectativas de grupos distintos:

- las diversas comunidades de usuarios del repositorios;
- otros sistemas, otras tecnologías con las que hay que asegurar una interoperabilidad que no siempre es sencilla.



Forma de evaluación/autoevaluación/revisión

Depende de la institución y sus objetivos, ¿qué se evalúa?

- ❑ Servicios, valor añadido.
- ❑ Objetivos cumplidos, grado de colaboración con otros repositorios, estado de desarrollo, personal...

Premisas:

- ❑ Es distinto evaluar un repositorio en sus inicios que cuando ya está desarrollado.
- ❑ No hay una única forma de mirar la calidad de un repositorio.



Alternativas de los indicadores

- Enfocados en aspectos tecnológicos.
- Enfocados en la promoción.
- Enfocados en los servicios.



Al finalizar se servirá un vino regional

Algunas consideraciones

El rendimiento del repositorio, pero ¿entendido en términos de qué?: por ejemplo la gestión del repositorio.

Al principio la evaluación se orientó a lograr aumentar contenidos, después se pensó en otros aspectos: tecnológicos, procesos, personal, económicos, difusión, formación de usuarios.

Al evolucionar los repositorios empiezan a aparecer otros requerimientos como el cumplimiento con la legislación de AA.



Antecedentes importantes

- Serrano, Melero y Abadal (2014) recogen 10 años de estudios de evaluación y destacan algunos indicadores:
- Westell (2006): políticas, apoyo institucional, modelo de financiación, digitalización retrospectiva, interoperabilidad, promoción, preservación...
- Swan (2012): contenidos, usuarios, flujos de trabajo y financiación.
- Kim, Kim (2008): contenido, gestión y políticas, sistema y usos&usuarios.
- Cassella (2010): indicadores del rendimiento internos y externos.

Antecedentes importantes

- Serrano, Melero y Abadal (2014) recogen 10 años de estudios de evaluación y destacan algunos indicadores:
- Guía ReBIUN, Fecyt y Recolecta (2010 y 2014): visibilidad, políticas, aspectos legales y metadatos, interoperabilidad, estadísticas y seguridad, autenticidad e integridad de los datos.
- La Encuesta del Primary Research Group (2012) definió un conjunto de criterios: promoción, acceso y estadísticas, economía, cooperación, impacto, catalogación, derechos financiación y contenido.
- Vierkant (2013): Censo sobre repositorios institucionales alemanes (8 indicadores).

Conclusiones de los autores

El trabajo de Serrano, Abadal y Melero culmina analizando las categorías repetidas en unos y otros análisis y realiza una propuesta propia.

La comparación de los indicadores propuestos en la bibliografía revisada han sido clasificados en las categorías: marketing, económicos, interoperabilidad y cooperación, políticas, contenidos (sólo aparece la propuesta de Swan) y servicios de valor añadido.

SERRANO VICENTE, Rocío; MELERO MELERO, Remedios; ABADAL, Ernest. Indicadores para la evaluación de repositorios institucionales de acceso abierto. *Anales de Documentación*, [S.l.], v. 17, n. 2, sep. 2014. ISSN 1697-7904. Disponible en: <<http://revistas.um.es/analesdoc/article/view/190821>>. Fecha de acceso: 24 oct. 2017 doi:<http://dx.doi.org/10.6018/analesdoc.17.2.190821>.

Propuesta de Serrano, Melero y Abadal

Consideran dos aspectos:

- **Factores técnicos o internos del repositorio y Factores institucionales.**

Establecen 32 indicadores separados en 5 categorías: tecnologías, procedimientos, contenidos, marketing y personal.

El trabajo cuenta con una segunda parte de recolección de datos en función de encuestas (a gestores de repositorios e investigadores) y revisión de contenidos, la totalidad del trabajo se consolida con la Tesis Doctoral de Rocío Serrano Vicente de este año 2017 titulada “Evaluación de los repositorios institucionales de Acceso Abierto en España”.



Repository Audit and Certification DSA-WDS Partnership WG Recommendations

[Home](#)

Este grupo de la Alianza está dedicado a crear procedimientos armonizados para la certificación de repositorios en un nivel básico. Establecieron 16 requisitos organizados en 5 criterios:

- Los datos se deben encontrar en Internet
- Los datos deben ser accesibles (llevar licencia y tener derechos explícitos).
- Los datos deben estar en un formato utilizable
- Los datos deben ser confiables.
- Los datos deben tener un identificador único y permanente.

Autoevaluación y revisión por pares para otorgar el sello.



Certificado DINI



DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

SUCHEN

DINI (2006) se funda en las universidades alemanas con el fin de mejorar los servicios de información. Crean un certificado que es aplicable sobre repositorios de AA con 8 dimensiones y dentro de cada una aspectos obligatorios (M) y otros recomendados (R):

- Visibilidad del servicio,
- políticas,
- soporte para autores y editores,
- aspectos legales,
- seguridad de la información,
- indexación e interfases,
- acceso a estadísticas y
- disponibilidad a largo plazo.



¿Cómo obtenerlo?

- ❑ Se paga un pequeño canon.
- ❑ El gestor del repositorio solicita la certificación DINI tras completar un formulario en línea: <https://dini.de/dini-zertifikat/fragebogen/>
- ❑ Una vez completado el formulario y enviada, la solicitud, los datos serán verificados - generalmente se nombrarán dos revisores.
- ❑ Debe darse acceso a los servicios a ser certificados a los revisores.
- ❑ Los responsables del servicio deben responder a las preguntas de los revisores.
- ❑ Para obtener el certificado no basta con cumplir los criterios obligatorios, hay que cumplir con los recomendados.

ISO 16363 antecedentes

El propósito principal de esta norma es definir una Práctica Recomendada en la que basar un proceso de auditoría y certificación para evaluar la confianza de los repositorios digitales.

El desarrollo del Modelo de Referencia de Sistema Abierto de Información de Archivo (OAIS) consensuó lo que se requiere para que un repositorio digital provea conservación a largo plazo.

Las instituciones comenzaron a declararse a sí mismas “Conforme a OAIS” para marcar lo confiable de sus repositorios. ¿Cómo justificar esto?

Se hizo vital establecer criterios capaces de dar cuenta si un repositorio digital es confiable o no. En 2002 RLG y OCLC establecieron los criterios TRAC.

ISO 16363

El documento sigue los criterios TRAC para generarse como norma. Las secciones 1 y 2 son informativas y las secciones 3-5 proporcionan las métricas:

Las que abarcan la infraestructura organizacional con 25 requisitos.

- Viabilidad de la institución, estructura organizativa y personal, sostenibilidad económica, contratos y licencias, etc.

Las que abarcan la Gestión de objetos digitales con 42 requisitos: sigue la 14721.

Las que abarcan la Gestión de Riesgos de Seguridad y de Estructura Infraestructura y gestión de seguridad con 23 requisitos:

- Riesgos de las infraestructuras físicas, gestión de seguridad, asignación de roles.

Algunos acuerdos afortunadamente ...

Interoperabilidad: nada sirve si los distintos sistemas no están interconectados.

La interoperabilidad es un factor siempre presente a la hora de medir calidad.

- Directrices OpenAire: para repositorios de literatura, de datos y sistemas CRIS.
- En nuestro país: **Directrices SNRD:** Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales (2013 y 2015).



Directrices SNRD

Directrices para proveedores de contenido
del Sistema Nacional de Repositorios Digitales

Ministerio de Ciencia, Tecnología e Innovación Productiva

Corolario: muchos aspectos, pero mucho en común

1. Contenidos
2. Catalogación y metadatos
3. Interoperabilidad
4. Visibilidad
5. Preservación
6. Servicios
7. Estadísticas
8. Seguridad y actualizaciones
9. Comunicación y marketing

Sobre el contenido un párrafo aparte

- Cantidad de materiales: ¡en acceso abierto y a texto completo!
- Tipos de licencia: cuestiones a reflexionar sobre el tipo de licencia CC más conveniente.
- Datos: ¿qué datos? ¿en qué sistemas?

Nuestra propuesta





Personal

Personal

- Presenta un equipo de trabajo multidisciplinario con conocimientos de catalogación, gestión, informática y comunicación.
- El personal se nuclea dentro del repositorio, es decir, su lugar de trabajo es en el repositorio.
- Se mantienen redes de colaboración y asistencia con otras instituciones.
- Prevalece la adopción de estándares por sobre las decisiones locales.
- Hay formación continua del personal.



Visibilidad, promoción y apertura

- Existe un enlace al repositorio desde el portal principal de la institución
- El repositorio está inscripto en los principales directorios de repositorios: OpenDoar, ROAR, OAI Data Providers, etcétera.
- El repositorio es cosechado/indexado por recolectores nacionales e internacionales: ej SNRD, BASE, Google Scholar, etcétera.
- La URL del repositorio (dominio + ruta) contiene el nombre propio del repositorio pero no la tecnología con la que está implementado. Ej: /xmlui, /jspui, /dspace, etc.
- Posee perfiles activos en las redes sociales: Facebook, Twitter, entre otros.
- Cuenta con un blog institucional y/o del personal, sobre el que se realizan actualizaciones periódicas.
- Realiza promociones de diverso tipo y dirigidas a distintos públicos: cursos, seminarios, conferencias, talleres, folletos.
- Se incentiva el autoarchivo.
- Se expone el porcentaje de recursos abiertos de acuerdo a su tipología documental.

Interoperabilidad

- Servidor OAI-PMH
 - Exposición de recursos en Dublin Core Element Set (DC simple).
 - Exposición de recursos en conformidad con las [directrices 2015 del SNRD](#).
 - Ofrece SETs lo que permite a los agregadores cosechar los recursos que interesan.
 - Mantiene registro de ítems borrados (borrado persistente).
- Exportación a otros formatos más elaborados: RDF, Refworks, JSON, etc..
- Soporta ingesta vía SWORD.
- Permite la recuperación de datos a partir de OpenSearch y/o Feeds RSS.
- Interopera de forma transparente con otros sistemas internos de la universidad: portales, portales de revistas, de congresos.



Funcionamiento global del repositorio

- Utiliza uno o más esquemas de metadatos estandarizados.
- El perfil de metadatos incluye metadatos técnicos y de preservación.
- Se utilizan vocabularios controlados para la descripción.
- Se dispone de un módulo de embargo para bloquear temporalmente la publicación de recursos.
- Asocia identificadores persistentes a todas las obras (handle, DOI u otros).
- El procedimiento de carga varía en función de los permisos del usuario.



Políticas y documentación

Cuenta con documentación sobre:

- Los objetivos del repositorio: misión, visión, etcétera.
- El procedimiento de autoarchivo y su finalidad.
- Políticas generales del repositorio: datos, metadatos, preservación, reúso...
- Aspectos legales: licencias, versiones de las obras.
- Plan estratégico de conservación que defina el enfoque a largo plazo.
- Plan de continuidad y contingencia.

Además, se debe validar que cada recurso posee:

- Una licencia de distribución (que incluye la responsabilidad del autor y la declaración de autoría) y una licencia de uso, preferentemente en un metadato específico.



Infraestructura Tecnológica



- El software que sostiene al repositorio
 - Es un proyecto de código abierto, no propietario.
 - Tiene una comunidad que da soporte.
 - Está desarrollado usando lenguajes y librerías ampliamente difundidas.
 - Recibe actualizaciones regulares.
- El *frontend* público
 - Ofrece una performance aceptable para los usuarios.
 - Es usable desde diversos dispositivos (ejemplo: móviles).
 - Tiene un módulo de búsqueda y exploración de documentos.
 - Está optimizado para ser navegado y analizado por motores de búsqueda (SEO).
 - Ofrece acceso a estadísticas públicas.

Recomendaciones generales

- Hardware:
 - Memorias con detección y corrección de errores (ECC RAM)
 - Sistema de alimentación ininterrumpida de energía (UPS)
 - Fuente de alimentación redundante
 - Esquemas redundantes de datos (raid 1+)
- Sistema operativo:
 - Filesystem robusto con mecanismos de comprobación y recuperación de errores.
 - Se aplican automáticamente actualizaciones de seguridad del fabricante
 - Regularmente se aplican otras actualizaciones disponibles
 - Se utiliza un Sistema Operativo para Servidores
- No usar el servidor para otros propósitos que no sean del repositorio.
- Configurar los recursos asignados a cada servicio en función del uso (memoria / buffers, cantidad de procesos y/o accesos simultáneos)

Seguridad

- Uso de reglas de firewall para controlar accesos en función de su origen, destino, puertos entrantes y salientes.
- Usar mecanismos de detección de ataques por fuerza bruta.
- Revisión periódica de usuarios con permisos de acceso al servidor.
- Revisión de ingresos autorizados (logs).
- Uso de conexiones seguras para administración remota (ej. VPN).
- Forzar uso de contraseñas con longitud y complejidad “aceptable” para evitar ataques por diccionario de palabras.

Servidor - Monitoreo

- Control recursos mínimos disponibles (Watchdog).
- Implementar un mecanismo de control de logs (ej. logrotate) que
 - rote y descarte logs grandes, antiguos y/o innecesarios,
 - rote y comprima logs de accesos.
- Monitorear automáticamente disponibilidad de servicios desde el exterior del repositorio: testear performance, uptime, conectividad, etcétera.
- Detección y bloqueo de IPs con comportamiento abusivo (demasiado frecuente o malintencionado).
- Realizar sumas de comprobación (checksum) y validación de ficheros.
- Revisar correcta ejecución de tareas programadas.

Resguardo de datos

- Utilizar versionado del código fuente y configuraciones.
- Generar backups de:
 - configuración del servidor: asignación de recursos, permisos, tareas programadas, etc.
 - software y sus configuraciones
 - archivos binarios / assetstore
 - logs de acceso
 - bases de datos
 - datos complementarios para funcionamiento de servicios (Solr)
 - registros estadísticos
 - sistemas complementarios (ej. de autoridades)
- Definir un esquema de rotación de backups en función de su importancia y frecuencia de cambios.
- En caso de usar un entorno de virtualización, realizar snapshots frecuentes e imágenes completas periódicamente.

¡Muchas gracias!

Por consultas: marisa.degiusti@sedici.unlp.edu.ar

Nuestros sitios

<http://sedici.unlp.edu.ar>

<http://digital.cic.gba.gob.ar/>

<http://cesgi.cic.gba.gob.ar/>

<http://prebi.unlp.edu.ar>

<http://www.istec.org/liblink/>

<http://revistas.unlp.edu.ar/cientificas/>

<http://revistas.unlp.edu.ar>

<http://congresos.unlp.edu.ar>

<http://ibros.unlp.edu.ar>