

Matomo como herramienta para la obtención de Estadísticas de Uso de los Documentos dispuestos en el Repositorio Institucional de Acceso Abierto de la Facultad de Tecnología y Ciencias Aplicadas, UNCA

Ivanna Maricruz Lazarte¹, María Vanesa Doria², Carola Victoria Flores², María Carolina Hausteín² y Efraín Córdoba³

¹ Departamento de Formación Básica, Facultad de Tecnología y Ciencias Aplicadas, UNCA

² Departamento de Informática, Facultad de Tecnología y Ciencias Aplicadas, UNCA

³ Instituto de Informática, Facultad de Tecnología y Ciencias Aplicadas, UNCA

{ilazarte, vanesadoria, carolaflores, caro-h,
efrainc}@tecno.unca.edu.ar

Resumen. Las estadísticas de uso permiten realizar un seguimiento de la producción científica depositada en los repositorios digitales de acceso abierto y estudiar las pautas de su crecimiento, ayudando de esta manera a diseñar estrategias futuras para alimentarlos con más contenido. Reflejan, además, la visibilidad, la difusión nacional e internacional y las tendencias de uso de los documentos dispuestos en abierto. También poseen una utilidad inmediata para autores, editores e instituciones para ayudar a justificar la importancia, relevancia, valor de su trabajo y complementar la información dada por las citas. Por lo expresado, resulta de gran importancia que los repositorios digitales cuenten con estadísticas de uso que permitan demostrar su utilidad, relevancia y, por consiguiente, sus beneficios. En este artículo se describe la implementación y el uso de Matomo en el Repositorio Institucional de Acceso Abierto de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca de la República Argentina. Matomo es una herramienta de Analítica Web que puede utilizarse para la obtención de estadísticas de uso de repositorios institucionales, con el propósito de facilitar el análisis y comprensión de los datos almacenados, medir su popularidad, utilidad y, además, contribuir a la correcta toma de decisiones.

Keywords: Analítica Web, Estadísticas de Uso, Repositorios Digitales, Acceso Abierto, Matomo.

1 Introducción

Un repositorio institucional es un conjunto de servicios que ofrece la universidad a los miembros de su comunidad para la gestión y diseminación de materiales digitales creados por la misma institución y sus miembros. Es esencialmente un compromiso organizacional a la custodia de estos materiales digitales, incluyendo su preservación a largo plazo, organización, acceso y distribución [1]. El material digital se compone de

archivos que representan una obra y metadatos que la describen. La producción científica publicada en los repositorios digitales abarca trabajos técnico-científicos, tesis académicas, artículos de revistas, entre otros, y suele ser resultado de la realización de actividades de investigación financiadas con fondos públicos ya sea, a través de sus investigadores, tecnólogos, docentes, becarios postdoctorales y estudiantes de maestría y doctorado [2].

Los repositorios digitales nacen del movimiento de Acceso Abierto, el cual fomenta que la producción científico-tecnológica se pueda, en forma gratuita, leer, descargar, reproducir, distribuir, imprimir, buscar o enlazar los textos completos de los artículos científicos, y usarlos con propósitos legítimos ligados a la investigación científica, a la educación o a la gestión de políticas públicas, sin otras barreras económicas, legales o técnicas que las que suponga Internet en sí misma. La única condición que plantea este movimiento para la reproducción y distribución de las producciones es otorgar a los autores el control sobre la integridad de su trabajo y el derecho a ser adecuadamente reconocidos y citados [2], [3], [4], [5], [6], [7].

Entre las ventajas de los repositorios digitales institucionales se pueden mencionar: maximizar la visibilidad, el uso y el impacto de la producción científica y académica en la comunidad internacional; retroalimentar la investigación; producir y/o dar soporte a las publicaciones electrónicas de la institución; facilitar el acceso a la información científica y académica. Entre las ventajas para los investigadores se puede mencionar: mayor rapidez en la publicación, mayor visibilidad, aumento de las citas y en consecuencia mayor impacto, centralización de la producción en un solo lugar, preservación a largo plazo. Para las instituciones las ventajas son: mayor visibilidad y prestigio, registro permanente de la actividad académica e investigadora, herramienta de marketing [3].

En nuestro país, en 2002 se creó la Biblioteca Electrónica de Ciencia y Tecnología [8], en 2008 se comenzó a impulsar el Acceso Abierto y en 2011 se creó el Sistema Nacional de Repositorios Digitales (SNRD) [9]. En 2013, se aprobó la Ley Nacional N° 26.899 [10] sobre la creación de repositorios digitales institucionales de acceso abierto, propios o compartidos. Esta Ley estableció, entre otras cosas, la obligatoriedad del acceso abierto a la producción financiada con fondos públicos a nivel nacional a través de repositorios digitales, transformando el tema en política pública. En 2016 se aprobó el Reglamento operativo para la aplicación de la mencionada Ley.

A medida que el Acceso Abierto, como modelo de comunicación científica, se arraigó en el mundo y en nuestro país, surgiendo más repositorios institucionales, se planteó la necesidad de encontrar criterios que permitan evaluar la evolución histórica del contenido y estado de los repositorios, la usabilidad de sus publicaciones científicas, o el uso de los servicios que ofrecen, entre otras cosas [2], [11]. Es por esto que, para demostrar el valor de los repositorios institucionales y garantizar su sostenibilidad en el tiempo, se requiere presentar evidencia objetiva que refleje el uso de los contenidos digitales dispuestos en abierto. Esa evidencia son los datos de uso que, por lo general, suelen ser presentados de forma agregada en estadísticas de uso [12]. Las estadísticas son una herramienta clave a la hora de medir un repositorio en aspectos como su crecimiento, la actividad de sus usuarios y el uso de su contenido, así también lo es la obtención de gráficos que resuman los datos calculados en algo más tangible; los

repositorios deben retroalimentarse con estos datos/información y utilizarlos bajo una política de expansión y mejora continua: sirven como control de calidad y permiten saber sobre el estado de avance en los repositorios [2].

La recolección sistemática de estadísticas puede ser una herramienta útil para que los repositorios puedan alcanzar objetivos internos y externos. El análisis del uso interno permite realizar un seguimiento de la producción científica depositada y estudiar así las pautas de crecimiento, ayudando a diseñar el plan de trabajo y las estrategias futuras para alimentarlo con más contenido. Reflejan también la visibilidad y la difusión internacional y las tendencias de uso de estos archivos abiertos, que son indicadores de su eventual consolidación. Por otra parte, las estadísticas pueden ser un medio persuasivo y elocuente para explicar el por qué son necesarios los repositorios abiertos ante la institución de la que dependen y su agencia financiadora –mostrando la relación costo-beneficio del repositorio– y ante la comunidad científica cuya investigación difunden y preservan –demostrando su efectividad en potenciar la accesibilidad de los resultados de investigación de un modo gratuito e inmediato en internet [10]. También, permiten a los investigadores saber cuánta atención está recibiendo su investigación y cómo los usuarios están accediendo a este material, comparando el grado de “popularidad” de sus trabajos con el de sus colegas [11].

En resumen, los análisis estadísticos son un valor añadido para los administradores de los repositorios y para sus usuarios ya que miden su popularidad y uso, y contribuyen a la correcta toma de decisiones, a fijar las prioridades y a elaborar mejores políticas científicas [11].

Por lo expresado, resulta de gran importancia que los repositorios digitales de acceso abierto cuenten con estadísticas de uso que permitan demostrar su utilidad y, por consiguiente, sus beneficios.

En un trabajo previo [13] se analizaron varias herramientas para la obtención de estadísticas de uso que puedan implementarse en el Repositorio Institucional de Acceso Abierto (RIAA) de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (UNCA) [14], [15]. En este trabajo se presenta la implementación de Matomo, una herramienta libre y de código abierto de Analítica Web, para la obtención de las estadísticas de uso del RIAA.

2 Marco Teórico

2.1 Estadísticas de uso

A medida que la cantidad de material en el repositorio crece en volumen y antigüedad, también crece su estructura, típicamente definida a partir de un conjunto de comunidades/colecciones, las interrelaciones entre sus objetos, el acceso y uso por parte del público. Debido al creciente volumen de datos en un repositorio, resulta vital para evaluar el estado del mismo a través de reportes y estadísticas que simplifiquen tanta complejidad.

Las estadísticas de uso indican directamente la actividad y el uso que un usuario del repositorio digital hace del sistema en sí y de la producción académica de la institución. Como puede verse en la Figura 1, la noción de uso puede ser interpretada a través de

transacciones de acceso donde cada transacción representa de forma unívoca un evento de tipo solicitud-respuesta entre un usuario (que solicita un recurso), y un servicio de información (que contesta el pedido). Por ejemplo, un evento de uso se produce cuando un usuario descarga un documento que se administra en un repositorio digital, o cuando un usuario ve los metadatos asociados con este documento. Cabe aclarar que la solicitud de un documento en sí mismo no dice nada sobre su uso real, es decir, si se usa para leerlo o citarlo. Así como el retiro de un documento de una biblioteca no supone necesariamente su lectura, la descarga de un documento de un repositorio tampoco la implica. No obstante, en ambos casos se supone que hay una intención latente.



Fig. 1. Evento de uso

Las estadísticas de uso se pueden clasificar en tres categorías: popularidad, comportamiento y descargas [16]. En la Tabla 1 se muestran algunos ejemplos de estadísticas de uso.

Tabla 1. Ejemplos de estadísticas de uso, según su categoría.

Ejemplo	Categoría
Cantidad de accesos a un recurso o colección	Popularidad
Documentos más descargados	Descargas
Cantidad de descargas por autor	Descargas
Cantidad de accesos a los recursos de una colección	Popularidad
Distribución de accesos por origen (continente/país/ciudad)	Popularidad
Tasa de accesos por fecha	Comportamiento
Fechas con mayor cantidad de accesos	Comportamiento

La disponibilidad de estadísticas de uso es, sin duda, un insumo informativo y valioso para la toma de decisiones respecto a múltiples aspectos asociados al funcionamiento, la promoción y la aceptación de un repositorio digital. La toma de decisiones se puede realizar en varios niveles [2]:

- **Político/estratégico:** las decisiones estratégicas son las que determinan las metas, los propósitos y la dirección de toda la institución y, por consiguiente, direccionan las metas del repositorio. Por ejemplo: cooperar con otros grupos que puedan ayudar

a fomentar ciertos contenidos institucionales, interactuar más con determinados actores de la institución, etc.

- **Táctico:** éstas decisiones se refieren al desarrollo de tácticas para cumplir las metas estratégicas que definieron los altos niveles administrativos, son más específicas y concretas que las estratégicas y más orientadas a las acciones. Por ejemplo: cambiar la forma de agrupar cierto tipo de recurso para aumentar su visibilidad, incorporar un nuevo tipo de recurso, implementar una nueva metodología de carga de publicaciones al repositorio, etc.
- **Operativo/tecnológico:** se refieren al curso de las operaciones diarias, estas decisiones determinan cómo se dirigen las operaciones diseñadas para cumplir con las decisiones tácticas. Por ejemplo: mejorar la estructura de determinada página del repositorio, conseguir más hardware y mejorar la conectividad al repositorio, ampliar el software que soporta al repositorio para integrar cierta tecnología, revisar los índices de la base de datos, etc.

Con respecto al beneficio para los gestores del repositorio, a través del análisis de las estadísticas de uso, se pueden realizar acciones concretas para solucionar problemas o realizar mejoras, por ejemplo: incorporar nuevos idiomas, a partir del origen de los usuarios [2].

Con respecto a los autores, estudios recientes han demostrado que, en una variedad de comunidades, las cifras de descargas de documentos de Acceso Abierto están fuertemente correlacionadas con las citas posteriores [17], [18]. Podría decirse que las estadísticas de uso representan una alternativa a las métricas basadas en citas, como el Factor de Impacto, porque tienen la ventaja de que se pueden recopilar en tiempo real, ya que se obtienen inmediatamente después de la publicación del documento. Los lectores son quienes determinan la métrica haciendo clic en los documentos individuales que les interesan. En cambio, las métricas basadas en citas solo se pueden medir con un retraso ya que las citas se cuentan retrospectivamente, es decir, después de que se haya publicado el artículo citado, y es esta cifra la que se utiliza para calcular el Factor de Impacto. Además de eso, solo las revistas enumeradas en el índice se incluyen en el cálculo [19]. En la Tabla 2 pueden verse las diferencias entre métricas basadas en citas y basadas en el uso.

Tabla 2. Diferencias entre métricas basada en citas y basadas en uso [20].

Métricas basadas en citas		Estadísticas de uso
Autor	¿Quién?	Lector
Retrasado (<i>delayed</i>)	¿Cuándo?	Inmediatamente
Revistas indexadas	¿Qué?	Cualquier objeto digital
Análisis a nivel de revista	¿Cómo?	Análisis a nivel de objeto

Las publicaciones de Acceso Abierto no están sujetas a ninguna restricción de acceso, lo que significa que son una base excelente para recopilar estadísticas de uso y para establecer un estándar comparable internacionalmente. Esto a su vez significa que, como forma de publicación, el Acceso Abierto representa una forma transparente y de bajo costo para evaluar la distribución y el uso de los resultados de las investigaciones

[19]. Es por ello que varias iniciativas a nivel internacional están trabajando en el desarrollo de estándares para la elaboración de estadísticas de uso, a través de una serie de normas uniformes que pauten las características que deben poseer los informes de estadísticas de uso para la recopilación, intercambio y análisis de los datos de uso de repositorios digitales [12]. Sin embargo, todavía no hay consenso sobre qué datos deben recopilarse, qué mecanismos de filtrado son apropiados y qué análisis son útiles para los académicos en diversas disciplinas [17].

2.2 Recolección de datos de uso

Existen diferentes técnicas para la recopilación de datos a utilizarse para la obtención de las estadísticas de uso, pero las más utilizadas son: archivos de registro (*logfile*) (ver Figura 2) y etiquetado de páginas (ver Figura 3) [20].

Los archivos de registro contienen la totalidad de las transacciones que se llevan a cabo entre los clientes (visitante) y el servidor web. Cada transacción que recibe el servidor web queda reflejada en una o varias entradas del archivo de registro. Pueden contener la dirección IP del cliente, la fecha y hora de la solicitud, el nombre del archivo solicitado, el estado y nombre de la respuesta HTTP, la URL de referencia y la información del navegador [12], [20]. El formato en que se registran estos datos se encuentra estandarizado, denominado *Common Log Format* (CLF), por lo que se pueden realizar procesos de parseo o conversión a base de datos con relativa facilidad.

La ventaja de esta técnica es que siempre es propietario de los datos recopilados para la obtención de las estadísticas y los archivos de registro son accesibles, por lo cual puede comenzar a crear informes básicos muy rápidamente, utilizando cualquier herramienta de análisis (*log file parsers*) [21]. La desventaja de esta técnica es que las páginas almacenadas en caché no son registradas en el archivo de log, porque no requieren una solicitud al servidor, y por lo tanto no son contabilizadas en las estadísticas. Además, como estos registros se crearon para capturar todos los éxitos en el servidor, se debe ser muy cuidadoso al eliminar datos indeseados como errores de página, tráfico de robots, etc., para obtener información precisa de tendencias y comportamiento [21].

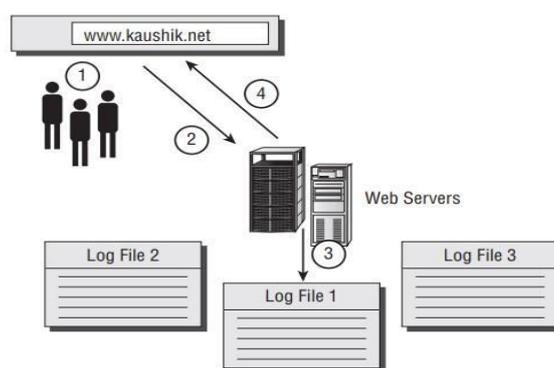


Fig. 2. Esquema simple de cómo los archivos de registro capturan los datos [21].

El etiquetado de páginas es una técnica muy sencilla y consiste en insertar pequeños fragmentos de código en cada página o en un elemento global de su sitio web. Habitualmente están escritos en JavaScript, pero pueden usarse otros lenguajes, como Java. Los datos son recolectados a través de esos códigos mientras se carga la página y son enviados a una base de datos. Una gran ventaja de esta técnica es que se tiene control sobre qué datos se recopilan y la variedad de datos que se pueden capturar es enorme, incluyendo los clicks y la posición del cursor, el seguimiento de los movimientos del ratón y las pulsaciones sobre el teclado, el tamaño de la ventana de los navegadores y los plug-ins instalados. Esta técnica es muy utilizada en Análítica Web. Por ejemplo, Google Analytics y Matomo usan esta técnica [12], [20], [22]. La desventaja de esta técnica es que, si el usuario tiene deshabilitado JavaScript en el navegador, la información no se recolecta [21].

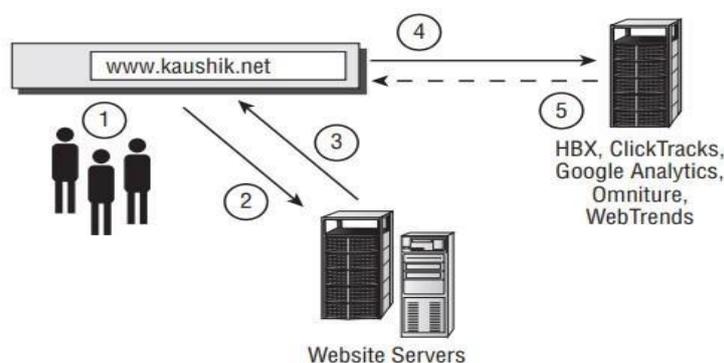


Fig. 3. Esquema simple de cómo funciona la captura de datos con el etiquetado de páginas [21].

Debido a que en el presente trabajo se utiliza una herramienta de Análítica Web para la obtención de estadísticas de uso de repositorios digitales, en la siguiente sección se describe brevemente la misma.

2.3 Análítica Web

La aparición de Internet y la creciente adopción de este medio como canal de comunicación y venta por parte del ambiente empresarial ha dado lugar al surgimiento de nuevas técnicas que analicen la eficacia y eficiencia de los sitios web [23]. La Análítica Web se centra en la medición, procesamiento, análisis y reporte del tráfico de Internet para entender y optimizar el uso de un sitio web [21]. Es una herramienta de negocio que ayuda a entender el comportamiento y la interacción de los usuarios en el sitio y a mejorar su experiencia de navegación.

Inicialmente su utilización se centró en el ámbito empresarial con el objetivo de analizar desde un punto de vista comercial los sitios web construidos por las empresas y su accesibilidad, pero últimamente su utilización se está acercando de forma progresiva al sector educativo, para diferentes aspectos como son el análisis del comportamiento de los usuarios en acciones formativas de e-learning, la comparación de las navegaciones realizadas en entornos de LMS y móviles, el comportamiento del

tutor, los tipos de navegaciones que realizan los alumnos o el análisis de las comunidades virtuales [23], [24]. También puede utilizarse para la obtención de estadísticas de uso de los repositorios digitales, como se describe en el presente artículo.

Independientemente de la técnica de recolección de datos y de la herramienta de Analítica Web que se utilice, la mayoría de los datos que se obtengan se basarán en los clics de los usuarios. Para entender más la Analítica Web, a continuación, se detallan brevemente algunos de los términos más utilizados [25]:

- **Conversión (*conversion*):** ocurre cuando el usuario completa una acción determinada.
- **Hít:** es la solicitud de un archivo del servidor web.
- **Página vista (*Page view*):** es el número de veces que ha sido visitada una página web.
- **Visita o Sesión (*visit/session*):** se define como el número de solicitudes de página realizada por el mismo usuario en un periodo de tiempo no superior a los 30 minutos entre cada petición de página.
- **Usuario Único (*Unique User/unique visitor*):** es el número de usuarios individuales que visitan un sitio en un periodo de tiempo determinado independientemente del número de veces que accedan a ese sitio.
- **Tasa de rebote (*Bounce Rate*):** es el porcentaje de visitas donde el usuario entra y sale de la misma página sin visitar otras páginas del mismo sitio.
- **Duración de la sesión (*Session Duration*):** es el tiempo promedio que los usuarios permanecen en el sitio cada vez que lo visitan.
- **Tiempo de actividad (*Active Time / Engagement time*):** es tiempo promedio que los usuarios pasan interactuando realmente con contenido en una página web.
- **Frequency / Session per Unique:** mide el número de veces que un usuario visita un sitio Web.

3 Obtención de estadísticas de uso del Repositorio Institucional mediante Matomo

El RIAA [14], [15] está desarrollado sobre la plataforma DSpace [26], el cual es un software de código abierto que permite gestionar repositorios de archivos (textuales, audio, vídeo, etc.), facilitando su depósito, organizándolos en comunidades y colecciones, asignándoles metadatos y permitiendo su difusión a recolectores o agregadores.

Si bien DSpace dispone del módulo DSpace Statistics encargado de la generación de algunos reportes estadísticos, el mismo presenta limitaciones en cuanto a las estadísticas que se pueden obtener [2], [12]. Este módulo genera reportes a partir de los accesos o visitas de páginas del repositorio, las descargas de bitstreams, las búsquedas en el repositorio y los eventos de workflow en el repositorio (ver Figura 4).

Es por ello que, para complementar estas estadísticas, se decidió usar otra herramienta que sea compatible con DSpace. Se analizaron las herramientas AWStats [27], Webalizer [28], Matomo [29] y Google Analytics [30] para la obtención de estadísticas de uso que puedan implementarse en el RIAA [13]. De estas herramientas,

se seleccionó Matomo debido a que es una herramienta libre y de código abierto, lo que implica evitar o reducir costos de actualización, soporte, licencias, etc. Esta decisión representa una gran ventaja para las universidades públicas que, en general, no cuentan con suficiente presupuesto para costear estos tipos de gastos. La idea es proporcionar una alternativa viable que, aprovechando al máximo las herramientas libres, amplíe el acceso a la información, reduciendo los costos y brindando rápido acceso a los resultados de la investigación.

Matomo es una herramienta de Analítica Web todo en uno, ya que ofrece una gama completa de funcionalidades que permiten evaluar todo el recorrido de los usuarios que visitan su sitio web.

Estadísticas
Número total de visitas

	Visualizaciones
Como trabajar Referencias y Bibliografía según Normas APA con Microsoft Word	359
Sitio web del Área de Educación a distancia de la Facultad de Tecnología y Ciencias Aplicadas	301
Analysis and Definitions of Efective Metrics for Scrum Teams	291
SISTEMA DE GESTIÓN ADMINISTRATIVA PARA LA ESCUELA SECUNDARIA N° 44 "JORGE LUIS BORGES"	271
Aplicación móvil de test de orientación vocacional para la Universidad Nacional de Catamarca como herramienta para ayudar a disminuir la deserción universitaria y promocionar las carreras	150
Directrices para la construcción de un repositorio temático	142
TIC en Educación: Implementación de videoconferencia en la plataforma Moodle de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca	120
Gestión de capital intelectual en el ámbito universitario sobre la función de investigación	116
Repositorios de Información como recurso de comunicación académica de Universidades Nacionales Argentinas	108
Tutorial de cómo Trasmittir por youtube desde 8x8	102

Fig. 4. Estadísticas de uso provistas por DSpace.

Matomo rastrea en tiempo real páginas vistas y visitas de un sitio web y muestra reportes de estos datos para su análisis. Algunas de sus características son las siguientes:

- Tiene una interfaz moderna, con un tablero configurable donde se pueden añadir y quitar widgets con arrastrar y soltar.
- Proporciona información sobre localización geográfica de los visitantes con mapas en tiempo real o de un período determinado.
- Muestra los perfiles de los visitantes, con información sobre cuándo fue la primera y última visita, dispositivos utilizados, navegador, sistema operativo, ubicación, páginas visitadas, descargas, tiempos.
- Permite fijar objetivos para las acciones que se desea para los visitantes, como visitar una página, para hacer un seguimiento del número de visitas que resultan en la realización del objetivo.
- Permite el envío de email con informes.
- Permite tener la propiedad del 100% de los datos y proteger la privacidad de sus usuarios.
- Permite mostrar su marca en todos los reportes.

Tiene la ventaja de que almacena todos los datos en su propio servidor, lo que permite al cliente mantener el control total de sus datos. Además, utiliza todos los datos para hacer los cálculos de las estadísticas, lo que le da mayor precisión comparado con otras herramientas, como Google Analytics.

Corre sobre un servidor web PHP/MySQL. Para utilizarlo se requiere instalar en el servidor web, pero también cuenta con una versión Cloud de pago.

Soporta más de 45 idiomas. Se integra fácilmente con más de 100 tecnologías que incluyen sistemas de gestión de contenido, tiendas de comercio electrónico, foros, frameworks y más. Entre ellas se encuentran WordPress, Magento, Joomla, Drupal, WooCommerce, Microsoft Sharepoint y Moodle.

Todas estas características han hecho que Matomo sea la herramienta elegida para complementar el módulo de estadísticas de uso al RIAA, de manera que permita facilitar el análisis y comprensión de los datos almacenados, medir su popularidad y utilidad, y contribuya a la correcta toma de decisiones.

3.1 Instalación y configuración de Matomo

Antes de comenzar la instalación de Matomo, hay que asegurarse de que el proveedor de alojamiento web cumpla con los requisitos mínimos para su ejecución: Servidor web como Apache, Nginx, IIS, etc.; PHP versión 5.5.9 o superior; MySQL versión 5.5 o superior, o MariaDB; Extensión PHP pdo y pdo_mysql, o la extensión mysqli (habilitado por defecto).

Si es necesario, se deben instalar los paquetes requeridos. Luego se procede a crear una base de datos Matomo en blanco. Enseguida se procede a instalar Matomo, y a continuación, se deben cambiar los permisos de la carpeta raíz.

Luego, se debe configurar el archivo de configuración `matomo.conf` de Apache2 para Matomo. En dicho archivo se debe reemplazar la línea resaltada con su propio nombre de dominio y ubicación raíz del directorio, como se muestra en el siguiente fragmento de código:

```
<VirtualHost *:80>
    ServerAdmin admin@example.com
    DocumentRoot /var/www/html/matomo
    ServerName example.com
    <Directory /var/www/html/matomo/>
        Options FollowSymlinks
        AllowOverride All
        Require all granted
    </Directory>
    ErrorLog ${APACHE_LOG_DIR}/error.log
    CustomLog ${APACHE_LOG_DIR}/access.log combined
</VirtualHost>
```

Luego se debe reiniciar el servidor Apache2 e iniciar el asistente de instalación de Matomo, donde se debe indicar la información de conexión de la base de datos y crear una cuenta de administrador, entre otros aspectos. En este proceso se obtiene el código de seguimiento (un script) para agregar al sitio web a rastrear. En este caso el RIAA, el

cual está implementado en DSpace, por lo cual sólo se debe agregar el script al archivo `jquery-ui-1.8.15.custom.min.js`.

Matomo también provee información de la ubicación de los visitantes. Por defecto estima la ubicación basado en la configuración del idioma del sistema operativo del visitante. Para obtener información precisa y completa de la ubicación de los visitantes, es necesario descargar la base de datos GeoLite City desde MaxMind, extraer este archivo y copiar `GeoLite2-City.mmdb` dentro del subdirectorío `misc` de Matomo.

3.2 Estadísticas de uso obtenidas

Una vez que Matomo se encuentra funcionando, al acceder al mismo se muestra la página principal con el tablero (ver Figura 5) que puede ser configurado por el administrador.

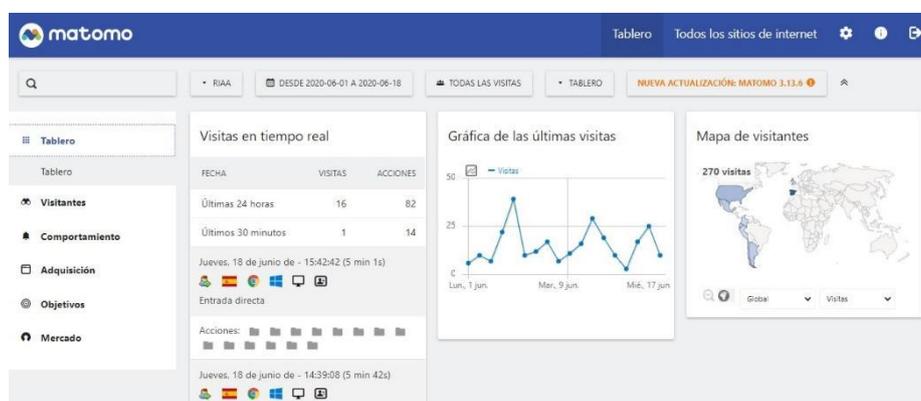


Fig. 5. Página principal de Matomo

A la izquierda de la página principal, se muestra el menú principal que permite generar múltiples reportes, organizados por tipo: Visitantes, Comportamiento, Adquisición, Objetivos, Mercado. Por ejemplo, el menú Visitantes permite generar los reportes General, Registro de visitas, Tiempo real, Mapa en Tiempo real, Ubicaciones, Dispositivos, Software, Tiempos, Ids de usuario y Variables personalizadas.

El reporte General del menú Visitantes (ver Figura 6), se pueden ver la gráfica de las últimas visitas, el cual es una visión holística de la evolución de las visitas y permite seleccionar las métricas a trazar (visitas, promedio de duración de las visitas, porcentaje de rebote, acciones por visita, visitas con conversiones, páginas vistas, descargas, búsquedas). En este reporte también muestra gráficos sobre la cantidad de visitas, duración media de la visita, % visitas que han rebotado (salieron del sitio después de haber visto una sola página), acciones del visitante (vistas de páginas, descargas, enlaces salientes y búsquedas internas dentro del sitio), tiempo medio de generación, páginas vistas, páginas vistas únicas, búsquedas totales en su sitio web, palabras claves únicas, cantidad de descargas, cantidad de enlaces de salida, cantidad máxima de acciones en una visita.



Fig. 6. Recorte del informe General, del menú Visitantes. Período 01/06/2020-18/06/2020.

En el menú Comportamiento, se pueden ver informes sobre páginas, páginas de entrada, páginas de salida, títulos de página, búsqueda interna, enlaces de salida, descargas (ver Figura 7), eventos, contenidos, compromiso y transiciones.

Descargas			
URL DE DESCARGA ▲	DESCARGAS ÚNICAS	DESCARGAS	
repositorios.tecno.unca.edu.ar:8080	126	137	
/bitstream/handle/RIAFTYCAUNCA/22/TESIS DORIA electronica.pdf?sequence=1&isAllowed=y	1	1	
/bitstream/handle/RIAFTYCAUNCA/23/Lazarte-WICC2019.pdf?sequence=1&isAllowed=y	4	4	
/bitstream/handle/RIAFTYCAUNCA/25/Repositorios digitales como recurso de comunicación.pdf?sequence=1&isAllo...	1	1	
/bitstream/handle/RIAFTYCAUNCA/28/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y	2	2	
/bitstream/handle/RIAFTYCAUNCA/31/Lazarte_Desarrollo-de-Software-RETICA.pdf?sequence=1&isAllowed=y	1	1	
/bitstream/handle/RIAFTYCAUNCA/37/Trabajo_Final_Carabús - Maubecin.pdf?sequence=1&isAllowed=y	1	1	
/bitstream/handle/RIAFTYCAUNCA/39/Trabajo_Final-Moreno-Trossero - Vergara - Febrero 2017 .pdf?sequence=1&isAll...	2	2	
/bitstream/handle/RIAFTYCAUNCA/40/Trabajo_Final-MERELES-RASGIDO 2017-07.pdf?sequence=1&isAllowed=y	1	1	

Fig. 7. Reporte de Descargas, del menú Comportamiento. Período 01/06/2020-18/06/2020.

4 Conclusiones

Los repositorios institucionales son clave para promover la visibilidad de los resultados de la investigación financiadas por el Estado. A medida que la cantidad de material en el repositorio crece en volumen y antigüedad, también crece su estructura, típicamente definida a partir de un conjunto de comunidades/colecciones, las interrelaciones entre sus objetos y el acceso y uso por parte del público.

Debido al creciente volumen de datos en un repositorio, resulta vital evaluar el estado del mismo a través de reportes y estadísticas que simplifiquen tanta complejidad. La

disponibilidad de estadísticas de uso es, sin duda, un insumo informativo y valioso para la toma de decisiones respecto a múltiples aspectos asociados al funcionamiento, la promoción y la aceptación de un repositorio digital. Su disponibilidad beneficia, en sus diferentes roles, tanto a los gestores de repositorios como a los autores que depositan en ellos sus obras. Con respecto al beneficio para los gestores del repositorio, a través del análisis de las estadísticas de uso, se pueden realizar acciones concretas para solucionar problemas o realizar mejoras. Con respecto a los autores, estudios recientes han demostrado que, en una variedad de comunidades, las cifras de descargas de documentos de acceso abierto están fuertemente correlacionadas con las citas posteriores.

Es por ello que se implementó el módulo de estadísticas de DSpace en el Repositorio Institucional de Acceso Abierto de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca y, debido a sus limitaciones, se complementaron las mismas con la herramienta de Analítica Web Matomo, de manera de proveer estadísticas de uso más completas del Repositorio para la realización de análisis más avanzados y complejos, que permitan demostrar su utilidad y beneficios, y contribuya a la Facultad en la correcta toma de decisiones en los niveles político/estratégico, táctico y operativo/tecnológico.

Como trabajo futuro queda la visualización de las estadísticas de uso de Matomo dentro de la interfaz de DSpace para que sean públicas para los usuarios del Repositorio.

Referencias

1. Lynch, C. A.: Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. *portal: Libraries and the Academy* 3(2), 327-336. Johns Hopkins University Press. ISSN: 1530-7131 (2003).
2. Adorno, F. G.: Reportes estadísticos para repositorios digitales desarrollados en DSpace. Trabajo Final de Grado. Universidad Nacional de La Plata (2018).
3. Bongiovani, P., & Nakano, S.: Acceso abierto en Argentina: la experiencia de articulación y coordinación institucional de los repositorios digitales en ciencia y tecnología. *e-colabora "Revista de ciencia, educación, innovación y cultura apoyadas por Redes de Tecnología Avanzada"*, 1(2), 163-179. ISSN: 2027-7415 (2011).
4. De Volder, C.: Los repositorios de acceso abierto en la Argentina situación actual. *Información, cultura y sociedad: revista del Instituto de Investigaciones Bibliotecológicas*, (19), 79-98. ISSN: 1514-327 (2008).
5. Budapest Open Access Initiative, <https://www.budapestopenaccessinitiative.org>, último acceso 19/06/2020.
6. Bethesda Statement on Open Access Publishing, <http://legacy.earlham.edu/~peters/fos/bethesda.htm>, último acceso 19/06/2020.
7. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, <https://openaccess.mpg.de/Berlin-Declaration>, último acceso 19/06/2020.
8. Biblioteca Electrónica de Ciencia y Tecnología, <https://www.biblioteca.mincyt.gob.ar/>, último acceso 19/08/2020.
9. Sistema Nacional de Repositorios Digitales, <https://repositoriosdigitales.mincyt.gob.ar/vufind/>, último acceso 19/08/2020.
10. Ley Nacional N° 26.899: *Repositorios digitales institucionales de acceso abierto*, <http://servicios.infoleg.gob.ar/infolegInternet/anexos/220000-224999/223459/norma.htm>, último acceso 19/08/2020.

11. Bernal, I., & Pemau-Alonso, J. Estadísticas para repositorios: sistema métrico de datos en Digital. CSIC. El profesional de la información, 19(5), 534–544 (2010).
12. Archuby, G., González, C. M., Jorquera Vidal, I., Merlino, C., Pichinini, M.: Medición de uso en repositorios digitales: Hacia la construcción de un marco de referencia argentino. En III Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología. Universidad Nacional de La Plata (2013).
13. Lazarte, I. M.; Doria, M. V.; Flores, C. V.; Cruz, J. R.; Almonacid, N. E.; Haustein, M. C.: Análisis de herramientas para la obtención de estadísticas de uso de los contenidos académicos dispuestos en Repositorios Institucional de Acceso Abierto de la Facultad de Tecnología y Ciencias Aplicadas, UNCA. En II Jornadas Internacionales de Estadística Aplicada, Salta, ISBN 978-987-633-552-2, (2019).
14. Doria, M. V., Haustein, M. C., Flores, C. V., Lazarte, I. M., Korzeniewski, M. I., Moreno, J. P., Arias, O. A., Santillan, P. J.: Movimiento de Acceso Abierto en la educación y la ciencia, desde la Facultad de Tecnología y Ciencias Aplicadas de la UNCA. En XIV Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA. San Miguel de Tucumán (2019).
15. Doria, M.V, Korzeniewski, M. I.: Gestión Repositorio de Acceso Abierto en la Facultad de Tecnología y Ciencias Aplicadas de la UNCA, II Workshop de Prácticas Educativas Abiertas, San Luis, Argentina (2020).
16. Aguillo, I.: Measuring the institution's footprint in the web. Library Hi Tech, 27(4), 540–556 (2009).
17. Carr, L., Brody, T., & Swan, A.: *Repository statistics: what do we want to know?*, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.684.8558&rep=rep1&type=pdf>, último acceso 19/06/2020.
18. Barrueco, J. M.: *Medición del uso e impacto de documentos distribuidos a través de repositorios institucionales*. BiD: Textos universitarios de biblioteconomía i documentació, 20(6), (2008).
19. DINI Working Group Electronic: Standardised Usage Statistics for Open Access Repositories and Publication Services, <https://edoc.hu-berlin.de/bitstream/handle/18452/2152/13-oas-en.pdf?sequence=1&isAllowed=y>, último acceso 19/06/2020.
20. Mellins-Cohen Tasha: Una guía (técnica) amigable a COUNTER., <https://www.projectcounter.org/wp-content/uploads/2016/08/Spanish-technical-guide.pdf>, último acceso 19/06/2020.
21. Kaushik, A.: *Web analytics: an hour a day*. John Wiley & Sons, (2007).
22. Rico Pérez, Y. E., & Montesi, M.: Visibilidad y uso de colecciones digitalizadas: propuesta de un indicador de visibilidad relativa. Ibersid, 7 (2013).
23. Graván, P. R., Almenara, J. C.: Análítica web de la comunidad virtual DIPRO2. 0. RELATEC: Revista Latinoamericana de Tecnología Educativa 12 (1), 35–50 (2013).
24. Almenara, J. C., Llorente Cejudo, M.C., Marín Díaz, V.: Las analíticas webs como elemento de apoyo al análisis de sitios web educativos. Revista científica electrónica de Educación y Comunicación en la Sociedad del Conocimiento 13 (2), (2014).
25. Rovira, S. P.: Analítica web, una nueva disciplina para un mundo digital. Cuadernos de comunicación Evoca. Analítica Web, vol. 2, pp. 5–9, (2010).
26. DSpace, Homepage, <https://duraspace.org/dspace>, último acceso 19/06/2020.
27. AWStats. Homepage, <https://awstats.sourceforge.io>, último acceso 19/06/2020.
28. Webalizer. Homepage, <http://www.webalizer.org>, último acceso 19/06/2020.
29. Matomo. Homepage, <https://matomo.org>, último acceso 19/06/2020.
30. Google Analytics. Homepage, <https://marketingplatform.google.com/intl/es/about/analytics>, último acceso 19/06/2020.