

Ensayo Científico



Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle

Potentialities and limits for the analysis of data from learning management systems. The case of Moodle

Agustín Zanotti*

Universidad Nacional de Córdoba, Argentina-CONICET/Universidad Nacional de Villa María, Argentina

azanotti@unvm.edu.ar

<https://orcid.org/0000-0002-7662-7593>

Lila Isabel Pagola

Nacional de Villa María, Argentina/Universidad Nacional de Córdoba, Argentina-CONICET

lpagola@unvm.edu.ar

<https://orcid.org/0000-0002-9933-8292>

Sección: **Ensayo científico**

Fecha de recepción: **18/08/2022** | Fecha de aceptación: **12/12/2022**

Referencia del artículo en estilo APA 7^a. edición:

Zanotti, A., & Pagola, L. I. (2020). Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle. *Transdigital*, 3(6), 1–23.

<https://doi.org/10.56162/transdigital145>



Licencia [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

International License (CC BY 4.0)

Resumen

El artículo reflexiona sobre el uso de informes, registros y bases de datos provenientes de sistemas de gestión del aprendizaje (en inglés, *Learning Management Systems* o LMS) para describir y analizar procesos educativos en universidades públicas de Argentina. Comienza con algunas discusiones referidas a la datificación y plataformización en el sector educación. Junto con ello, se incorporan definiciones sobre métodos digitales para las humanidades y ciencias sociales, así como analíticas de aprendizaje. Tales aproximaciones nos permiten considerar los datos generados por LMS con objetivos tanto pedagógicos como de investigación. Nos enfocamos en el caso de *Moodle*, un sistema de código abierto de los más extendidos en la actualidad, desarrollado por una comunidad de firmas, administradores, educadores y usuarios. En un primer momento, observamos las condiciones de acceso a los datos y el tratamiento de la información que poseen. En un segundo momento, analizamos un informe de registro de actividad para avanzar sobre tres niveles de lectura e interpretación: denotativo, connotativo y deconstructivo. Los resultados comparten algunas valoraciones sobre las potencialidades y límites de estos recursos, enmarcando los hallazgos dentro de un campo de indagación emergente e innovador. La investigación se basa en observaciones sobre plataformas y datos LMS, entrevistas con especialistas, fuentes secundarias, documentación de proyectos y recursos disponibles en línea.

Palabras clave: tecnologías de la información y de la comunicación; política de la información; información pedagógica; base de datos; tratamiento de la información

Abstract

The article reflects on the use of reports, records, and databases from Learning Management Systems (LMS) to describe and analyze educational processes in public universities in Argentina. It begins with some discussions regarding datafication and platformization in the education sector. Along with this, descriptions of digital methods for the humanities and social sciences, as well as learning analytics, are incorporated. Such approaches allow considering the data generated by LMS for both pedagogical and research purposes. We focus on the case of *Moodle*, one of the most widespread open-source systems today, developed by a community of firms, administrators, educators, and users. At first, we observe the conditions of access to the data and the treatment of the information they possess. In a second moment, we analyze an activity log report to advance on three levels of reading and interpretation: denotative, connotative, and deconstructive. The results share some assessments about the potentialities and limits of these resources, framing the findings within an emerging and innovative field of inquiry. The research is based on observations about LMS platforms and data, interviews with specialists, secondary sources, project documentation and resources available online.

Keywords: Information and communication technologies; information policy; pedagogical information; database; information processing

1. Introducción

El artículo reflexiona sobre el uso de informes, registros y bases de datos provenientes de sistemas de gestión del aprendizaje (en inglés, *Learning Management System* o LMS) para describir y analizar procesos educativos en universidades públicas de Argentina. Los LMS median diversos aspectos de los procesos de enseñanza-aprendizaje, permitiendo realizar actividades del curso, comunicar a estudiantes y profesores, así como evaluar aprendizajes. En la actualidad existen una variedad de alternativas, entre sistemas de software propietario y abierto.

Nos centramos en el caso de *Moodle*, un sistema de código abierto iniciado en 1999, en Australia, que tuvo su primera versión en 2002. Actualmente, cuenta con 300 millones de usuarios y 40 millones de cursos en 42 idiomas. Es desarrollado por la firma *Moodle Pty Ltd.* junto a una comunidad de desarrolladores, administradores y educadores. Se acompaña de una red global de *partners* que brindan soporte (Moodle, 2022a) y una asociación de usuarios sin fines de lucro que incide y moviliza recursos (Moodle Users Association, 2022).

La utilización de LMS genera conjuntos de datos, a los que se puede acceder por diferentes vías. La interpretación de tales datos, en tanto representación de la actividad de los usuarios/as, varía en función de las modalidades presenciales, híbridas o a distancia. Mientras en un contexto presencial los LMS se plantean como un recurso de apoyo que complementa la cursada (Jenaro-Río et al., 2018), en el escenario de la educación a distancia los registros de LMS ofrecen una oportunidad mayor para comprender los procesos de enseñanza y aprendizaje.

El interés por estos abordajes se vio incrementado con la pandemia de COVID-19, donde el confinamiento y las pedagogías de excepción otorgaron a las plataformas una nueva centralidad. Para reconstruir los procesos educativos desarrollados en estos últimos años, la consideración de los recursos diseñados por docentes y tutores, así como los

registros disponibles de las interacciones, pueden constituir fuentes de datos valiosas. Sin embargo, aún son pocas las experiencias sobre estas posibilidades. Mejorar la comprensión sobre las prácticas de implementación de LMS y sus posibilidades para la investigación, redundará en un beneficio para las comunidades educativas y una mayor eficacia de sus procesos.

El trabajo fue elaborado en el marco del proyecto *Recursos educativos abiertos: problemáticas de diseño en torno a usos y apropiaciones en entornos virtuales de aprendizaje*, radicado en la Universidad Nacional de Villa María, Argentina. Bajo una modalidad de investigación-acción, se buscó explorar las potencialidades de uso de conjuntos de datos extraídos de diferentes LMS utilizados en Universidades públicas provinciales y nacionales de la provincia de Córdoba, Argentina. Como segundo momento, se diseñaron recursos educativos abiertos y se realizaron capacitaciones.

Las primeras exploraciones consistieron en observaciones sobre las plataformas y entrevistas con especialistas en tecnología educativa. Permitieron reconstruir un panorama diverso, donde la sistematización de experiencias y buenas prácticas aparecía como una necesidad. Sobre esta base, el artículo organiza algunos primeros hallazgos e impresiones. Se basa en la reconstrucción y articulación del estado de arte sobre investigación basada en datos y educación, fuentes secundarias, documentación de proyectos LMS y recursos disponibles en línea.

La presentación comienza con discusiones referidas a procesos de datificación y plataformización en el sector de educación. Junto con ello, se incorporan definiciones sobre métodos digitales para las humanidades y ciencias sociales, así como analíticas de aprendizaje. Tales aproximaciones nos permiten considerar los datos generados por LMS tanto con objetivos pedagógicos como de investigación. Observamos en primer lugar las condiciones de acceso a los datos y el tratamiento de la información que poseen. En segundo lugar, avanzamos sobre diferentes niveles de lectura e interpretación que podrían ofrecer estos recursos. Hacia el final se comparten algunas valoraciones sobre sus potencialidades

y límites, intentando enmarcar los resultados en un campo de indagación emergente e innovador.

1.1 Plataformas digitales y educación

Los métodos digitales han sido definidos como el uso de tecnologías digitales para recolectar y analizar datos de investigación (Snee et al., 2016). Rogers (2013) distingue entre aquellos *digitalizados*, convertidos para su procesamiento y análisis, y aquellos *nativamente digitales*, creados en formato digital.

Los métodos digitales vienen a enriquecer los repertorios de las disciplinas sociales y humanas. Al tiempo que complementan las metodologías convencionales, favorecen la interdisciplina y los nuevos perfiles profesionales. Se trata de un espacio en construcción, en cierto sentido *precario*, donde la transformación de las plataformas y hábitos de los usuarios se combina con la exploración permanente de nuevos abordajes (Bruns y Burgess, 2016). Se plantea, asimismo, la necesidad de mejorar las habilidades técnicas asociadas a la extracción, resguardo, análisis y visualización de datos, lo cual requiere habilidades específicas no muy difundidas entre los científicos sociales (Manovich, 2011).

Según Srnicek, las plataformas son infraestructuras digitales que permiten que dos o más grupos interactúen, posicionándose como intermediarios que reúnen a diferentes usuarios (2018, p. 55). Van Dijck (2016) piensa a las plataformas digitales como ensamblajes sociotécnicos, estructuras performativas que moldean la socialidad *online*. Estas tienen capas visibles (interfaces, *defaults*) y ocultas (algoritmos, metadatos, protocolos). Sus formas de propiedad, modelos de negocio y formas de gobierno, son traducidas en políticas y condiciones de servicio en constante negociación. Sus modelos comerciales dominantes incluyen la concentración de gran cantidad de datos de contenido y de usuario, generados por hardware y software (Van Dijck, Poell y De Waal, 2018). Las plataformas digitales pueden ser con o sin fines comerciales. En este último caso son desarrolladas por una comunidad,

que suele ser liderada por una fundación sin fines de lucro o una firma basada en modelos abiertos (Vélez y Zanotti, 2020).

Las tendencias a la digitalización y plataformización de la vida social se hacen presentes en casi todos los dominios (Poell, T., Nieborg, D., & Van Dijck, J.; 2019; Zukerfeld, 2020). Una de las aproximaciones que habilitan las plataformas digitales es el análisis de las interacciones de los usuarios. Estos contienen tanto datos como metadatos (marcas de tiempo, información de emisores y receptores, etiquetas, entre otros). Los metadatos son *datos acerca de los datos* y sirven para suministrar información sobre los datos producidos (Geoidep, 2017). Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características de los datos (Geoidep, 2017). Tales registros fueron creados para fines específicos, sin la mediación del investigador, y necesitan de valoraciones metodológicas para tornarse significativos (Lomborg, 2012). Junto con ello, la preservación del anonimato y la privacidad adquieren centralidad. La evolución de los entornos tecnológicos hace que las consideraciones éticas sobre investigación tengan que ser frecuentemente actualizadas (Franzke et al., 2020).

Ardévol (2016) sostiene que las investigaciones basadas en datos tienen una epistemología abductiva, más inductiva que deductiva. Se parecen en este sentido a la *grounded theory* o la etnografía, coincidiendo en un enfoque hermenéutico. Del otro lado, Rogers propone que la comprensión hermenéutica va dejando lugar a un reconocimiento de patrones de comportamiento social. En este punto, señala, sin embargo, una diferencia importante entre la investigación de mercado y la científico-académica (DMI, 2015). Asimismo, Sued (2019) mantiene que es necesaria una traducción epistemológica norte-sur de los programas sobre métodos digitales, incorporando elementos críticos. Coincide en este sentido con Kitchin (2014), quien pondera los riesgos del positivismo y el empirismo en la investigación social.

Durante la última década, los servicios educativos en línea se ampliaron. La pandemia de COVID-19 vino a profundizar esta tendencia. Ello suscitó un interés por el análisis de datos

de LMS. Entre sus promotores, se sostiene que podría utilizarse para determinar el tiempo que un estudiante necesita para resolver un problema, las etapas cognitivas involucradas, la cantidad de instrucción necesaria o las interacciones, con el objetivo de optimizar los aprendizajes y adaptarlos a cada particularidad (Van Dijck, Poell y De Waal, 2018). Los críticos han señalado, sin embargo, que un énfasis en la datificación puede ser ineficaz en términos de impulsar los rendimientos o niveles académicos, al tiempo que se intensifican el gerencialismo y la vigilancia de estudiantes y docentes (Selwyn, 2016; Thompson, 2017, en Van Dijck, Poell y De Waal, 2018).

En educación, se plantea un debate en torno a la introducción de analíticas de aprendizaje predictivas (Ferrero, 2018). Retomando el caso de *Moodle*, la incorporación de tecnologías analíticas comenzó en 2015 y sus primeros modelos fueron implementados a partir de *Moodle* 3.4. Moodle solo incluye modelos que han demostrado ser buenos para predecir en una amplia gama de sitios y cursos, probados en conjuntos de datos anónimos de instituciones colaboradoras. En la actualidad cuenta con dos modelos integrados: Estudiantes en riesgo de deserción y Sin enseñanza (Moodle Docs, 2022b). A diferencia de otras plataformas comerciales, la captación de datos no está vinculada en este caso a un modelo de negocios, sino que es una consecuencia de su funcionamiento.

La literatura sobre *Moodle* se extiende entre varias disciplinas, principalmente centrada en su adopción, métodos para mejorar el aprendizaje, así como tendencias para la investigación y el desarrollo (Gamage, Ayres y Behrend, 2022). Recientemente, comienza a aparecer un interés por las posibilidades de *Moodle* en cuanto fuente de datos. Existen escasos antecedentes, referidos al análisis de informes a nivel de cursos (Meza Fernández y Marquet, 2017; Jenaro-Río et al., 2018) o registros de múltiples cursos en un centro universitario (López-Zambrano, Lara y Romero, 2020). Tales estudios buscan reconocer patrones que puedan relacionarse con estilos de aprendizaje o modelos predictivos sobre el rendimiento de los estudiantes. Otros abordajes se refieren al tratamiento de la información personal y la privacidad, en cumplimiento con normativa vigente en regulaciones nacionales o regionales (Amo et al., 2020). También se ha documentado el desarrollo de complementos

de Moodle que incursionan en técnicas de inteligencia artificial para inferir estilos de aprendizaje y tomar acciones consecuentes (Campo, Amandi y Biset, 2021). El sistema de Moodle está diseñado para permitir la adición de plugins, pequeños programas complementarios que amplían un conjunto variado de funcionalidades y pueden ser desarrollados por administradores (Moodle, 2022b).

2. Desarrollo

2.1 Desafíos en el acceso a los datos

Como anticipamos, Moodle genera registros y conjuntos de datos (*datasets*), que pueden ser utilizados tanto con objetivos pedagógicos como de investigación. Veamos cada tipo de recurso con mayor detalle:

- La base de datos “en bruto” de Moodle se compone de 429 tablas, donde las dos de mayor complejidad son “user” (información de usuario/a) y “course” (información de aula). La primera tiene 355 tablas dependientes -denominadas “children”- y la segunda tiene 154 (Examulator, 2022). Esta base está diseñada para el funcionamiento de la plataforma y contiene información sobre los cursos, sus componentes y recursos (mensajes, tareas, módulos, etc.), así como sobre los/as usuarios/as (perfil, inicio y última actividad registrada, cursos y actividades en los que participa, roles, etc.).
- Los registros o “logs” son los asientos de acciones realizadas en la plataforma por los usuarios. Incluyen cada interacción de usuarios/as con el LMS: cada vez que inicia sesión, cada clic, cada documento leído, cuánto tiempo, cada palabra escrita, entre otras. Todos los datos y metadatos se registran (Amo et al., 2020).
- Los reportes son consultas SQL que los administradores pueden realizar a la base de datos a través de una interfaz. Un listado de consultas, construido por la comunidad, está disponible en el sitio de Moodle Docs (2022a). Otros LMS no ofrecen esta posibilidad.

- *Informes de actividad* de cursos, a los que pueden acceder los usuarios/as con perfil docente. Siguiendo a Jenaro-Río et al. (2018), *Moodle* distingue entre: (1) acceso a tareas, estado de la entrega y valoración, (2) acceso al sistema; (3) acceso a recursos; (4) acceso a encuestas; (5) acceso a foros; (6) envío de archivos.
- *Informes generales* de todo el sitio, con fines de gestión y mantenimiento.

Antes de poder realizar cualquier análisis, es necesario el acceso a los datos. En las instancias de Moodle relevadas, los *datasets* y registros residen localmente, en servidores propios de las universidades donde se utiliza la plataforma. Ello marca una diferencia con otros LMS donde los datos son almacenados de manera remota e inaccesibles para usuarios o investigadores.

A pesar de estar disponibles localmente, observamos que no se seguían una serie de buenas prácticas para su posterior utilización. Una de ellas es el resguardo. Las áreas encargadas de la gestión del LMS resolvían borrar periódicamente los registros, variando entre los 30 días y un año. Ello en razón de que el volumen de información registrada tendía a crecer, lo cual generaba dificultades para su almacenamiento, procesamiento y recuperación.

Otra buena práctica no seguida es la duplicación de aulas. Ella debería ser una política académica en cada nuevo ciclo lectivo, manteniendo las aulas del periodo anterior. En su lugar, algunas universidades reiniciaban anualmente las aulas, eliminando las producciones e interacciones generadas, o superponiendo varias cohortes en un mismo espacio.

Junto con ello, las instancias de *Moodle* revisadas no implementaban términos y condiciones de uso. Esta funcionalidad fue incorporada a partir de la versión 3.4, en 2018. Se trata de una serie de cláusulas definidas por cada institución, que incluyen políticas de privacidad y acceso a los datos. Una vez que aceptan estas condiciones, los estudiantes pueden ejercer los derechos que otorga su legislación respectiva, como lo es el Reglamento

General de Protección de Datos (RGPD) en la Unión Europea (Amo, et al., 2020) o la Ley 25.326 de Protección de Datos Personales de Argentina.

Todo lo anterior da cuenta de que los fines de investigación no son imaginados ni contemplados por quienes realizan la gestión de los LMS, ni se utilizan asiduamente los informes en la actividad docente. Frente al potencial de la información registrada por los LMS, surge otra cuestión fundamental: la elaboración de un Plan de Gestión de Datos. El mismo contiene una serie de políticas que aseguran los usos y resguardos contemplados. Algunos aspectos aquí son (adaptado de CEPAL, 2020):

- Qué datos de investigación va a crear o recopilar.
- Quién será responsable de cada aspecto del plan de gestión.
- Cómo se organizarán los datos (estructuras de carpetas, convenciones de nomenclatura de archivos, versionamiento de archivos).
- Cómo se realizará la documentación de los datos durante la fase de recopilación y análisis de la investigación.
- Qué prácticas de administración se usarán para almacenar y proteger los datos (copias de seguridad, almacenamiento, control de acceso, archivado o destrucción cuando sea requerido).
- Qué instalaciones y equipos serán necesarios (espacio en disco duro, servidor de copia de seguridad, repositorio).
- Quién tendrá los derechos de propiedad y acceso a los datos.
- Cómo se conservarán y se pondrán a disposición los datos.

Una definición principal es si los recursos generados se considerarán como datos abiertos, “utilizados, reutilizados y redistribuidos libremente por cualquier persona” (OKI, s.f.). Junto con ello, se deberán establecer las adecuaciones necesarias para su puesta en disponibilidad. Existen técnicas para transformar datos confidenciales o sensibles previamente a que puedan ser compartidos. Entre ellas (adaptado de CEPAL, 2020):

- Agregación: la combinación de categorías relacionadas para proporcionar información a un nivel más amplio al que se toman observaciones detalladas
- Anonimización: se eliminan de los casos los datos identificadores como nombre y dirección.
- Seudonimización: una técnica común para proteger identidades en datos cualitativos.

Retomando el caso de *Moodle*, existen desarrollos para realizar la anonimización de los usuarios, de los registros y exportación de bases de datos (Amo et al., 2020; Kuo, Krahn, y Chang, 2021). La falta de definición sobre este conjunto de temas, ocasiona que los responsables tengan que decidir al momento de su requerimiento para fines de investigación.

2.2 Desafíos en la interpretación de los datos

Tomando como ejemplo un informe de actividad de curso, analizamos a continuación las posibilidades y límites de interpretación de los datos presentados. El *registro de actividad* es un tipo de informe *preseteado* de *Moodle* disponible para el rol docente. Contiene nueve columnas de datos (Tabla 1):

Tabla 1

Ejemplo de dataset creado por el informe "Registro de actividad" de Moodle

Hora	Nombre completo del usuario	Usuario afectado	Contexto del evento	Componente	Nombre evento	Descripción	Origen	Dirección IP
22/03/2022 15:14	Apellido y Nombre	-	Página: Recursos Audiovisuales y Transmedia para estudiar	Página	Módulo de curso visto	The user with id '18085' viewed the 'page' activity with course module id '128083'	web	000.210.68 .000

Poirier (2021) propone tres niveles de lectura, que seguiremos aquí, para el análisis de *datasets*: denotativa, connotativa y deconstructiva. Una lectura denotativa es literal del significado de los datos, tiene como objetivo discernir *lo que cuenta* según los productores de datos. Una lectura connotativa interpreta los intereses de sus creadores y partes interesadas, y las especificidades de los contextos culturales y geográficos de su producción. Finalmente, una lectura deconstructiva localiza los significados ausentes y las tensiones no reconocidas que siempre rondan las representaciones basadas en datos.

El primer nivel de lectura denotativa nos permite identificar qué información de *Moodle* es seleccionada para este informe. Analizando el *registro de actividad*, vemos que incluye datos de las personas involucradas en las acciones (*usuario* y *usuario afectado*, cuando un cambio modifica el estado de otro usuario, como cuando un docente califica), el *contexto* del evento, referido al elemento del curso involucrado, *componente* al nivel que contiene el elemento (la plataforma, un recurso o actividad específica).

También se incluyen *nombre evento*, una descripción sucinta de la acción para el analista (lo que representa un primer nivel de interpretación), y *descripción*, donde encontramos el log que se genera en la base de datos, por ejemplo: “*The user with id '271' viewed the section number '3' of the course with id '18660'.*” Por su parte “origen” indica si el evento se generó desde el sistema, desde la versión web o desde móvil. *Dirección IP* identifica el dispositivo digital que realizó la acción. Se trata de una etiqueta numérica que identifica la interfaz en la red de un dispositivo (computadora, laptop, teléfono inteligente) que utilice el Protocolo de Internet (en inglés, *Internet Protocol*, IP).

En este nivel denotativo se observa una cierta redundancia de datos, que funcionan como *traducciones* para facilitar la lectura del destinatario/a del informe. También observamos que los datos describen la supuesta acción que dejó el registro en términos muy generales (Tabla 2):

Tabla 2

Ejemplos de varios tipos de registros de las columnas "componente" y "Nombre evento" del informe "Registro de actividad" de Moodle

Componente	Nombre evento
Recurso	Módulo de curso visto
Sistema	Curso visto
Sistema	Evento de calendario actualizado
Tarea	Se ha visualizado el estado de la entrega.
Tarea	Retroalimentación vista

Por ejemplo, *curso visto* corresponde al acceso del curso por parte del usuario. Algo similar sucede con los módulos (todo recurso o actividad en la plataforma es un módulo). En algunos casos se describe con mayor precisión (tarea), en otros de forma genérica (y ciertamente poco útil, salvo que se lea con los datos de *descripción* donde figura el *id* del recurso). A modo de ejemplo, "Se ha visualizado el estado de la entrega" corresponde al registro descrito como: "The user with id '2400' has viewed the submission status page for the assignment with course module id '128054'. En este sentido, es posible observar los límites de estos informes que recogen datos de distintas tablas.

El segundo nivel connotativo permite preguntarnos sobre el significado de *actividad*. Esta denominación establece una selección de los datos que provienen de diversas tablas de *Moodle* (mientras que descarta otros) en relación a correlaciones posibles entre dicha actividad y procesos de enseñanza-aprendizaje.

En el ejemplo previamente citado, existe un nivel connotativo aplicado a esa descripción en la medida que se asume que el usuario que realizó una acción: vio, leyó, buscó o confirmó un dato, por ejemplo, la calificación o la devolución docente. Sin embargo, ese registro es básicamente una interacción con la plataforma por parte del usuario. Se están

tipificando entonces acciones frecuentes en situaciones comunes, que se asumen por parte del/la usuario/a a partir de indicadores basados en registros de la base de datos.

En esta línea, la recuperación del dato *Dirección IP* se relaciona con un uso posible de este informe: detectar actividades *anómalas* que impliquen posibles fraudes en casos de exámenes. La IP del/la usuario/a puede aportar precisiones a ese tipo de preguntas. Más allá de tales usos de verificación, cabe reflexionar: ¿qué preguntas pedagógicamente relevantes pueden hacerse con estos datos? ¿qué otra información se requerirá para otorgarles sentido?

Como anticipamos, la modalidad de enseñanza es un primer elemento clave. En la modalidad a distancia, cuando la plataforma es una mediación central de la propuesta de enseñanza, la actividad será potencialmente más representativa de los procesos en curso. Pero, en cualquier caso, estos registros deben ser interpretados en su contexto. La referencia de los/as protagonistas de la actividad (los/as estudiantes), permitiría comprender formas no previstas de usar las herramientas de la plataforma.

Un/a docente, por ejemplo, podría valerse del registro de actividad para comprender cómo sus estudiantes se preparan para una evaluación a partir de un incremento en los accesos a recursos de estudio. Tal como sostienen investigaciones precedentes (Jenaro-Río et al., 2018; López-Zambrano, Lara y Romero, 2020), serían esperables correlaciones del tipo “estudiantes que aprobaron son estudiantes que revisaron los recursos para estudiar, y otras orientaciones del/la docente previas al examen”, y también probablemente otras más complejas de interpretar, como cantidad de veces que se revisó un recurso, fechas de acceso, si el/la estudiante realizó consultas, actividades prácticas, actividades no obligatorias, entre otras.

La conformación -a partir de interacciones y acceso a recursos- de clasificaciones como *estilos de aprendizaje* (Meza Fernández y Marquet, 2017) son difícilmente generalizables. Los recursos usados por los/as estudiantes dejan diferentes registros de

datos, pero no tienen una correlación cognitiva evidente con los procesos de aprendizaje que habilitan.

Por ejemplo, un recurso en formato PDF sólo genera registros de acceso. El docente o investigador/a podrá conocer si el recurso fue accedido por su destinatario. Sin embargo, acceder a un archivo no implica estudiarlo, ni comprenderlo. Y quizá lo más relevante: no tener registro de acceso no significa estrictamente nada certero, en tanto que dicho recurso pudo ser accedido por otra vía, externa al aula.

Por el contrario, el recurso *libro* de Moodle, que permite construir clases interactivas originales en línea, genera registros por cada página accedida del libro. Los registros indican horario de acceso a cada una de sus partes, lo cual podría usarse para inferir el tiempo de permanencia del usuario/a. Ese dato podría tener un valor más preciso, siempre que no olvidemos que el tiempo de permanencia no significa tiempo de atención del usuario/a, especialmente en contextos multitarea.

Asimismo, otros recursos como cuestionarios, foros, *wikis* o actividades en las cuales los/as estudiantes producen contenido e interactúan, generan registros que pueden ser analizados complementariamente con metodologías cualitativas.

Por último, en el nivel de análisis deconstructivo, el *dataset* generado por el registro de actividad de Moodle nos permite pensar cuáles datos no fueron incluidos o no existen siquiera en la base de datos. Si analizamos la tabla correspondiente a *user* encontramos datos personales con varios niveles de diferenciación (nombre real, nombre de usuario, entre otros) (Examulator, 2022). En el diseño de esta tabla es posible notar que algunos descriptores responden a necesidades de las instituciones de enseñanza y no al funcionamiento del LMS. Un único identificador como número de identificación personal o legajo sería suficiente para distinguir a los usuarios. Sin embargo, tenemos campos para nombre, apellido y usuario (que se usa con el identificador único funcional). Algunas instancias de Moodle tendrán campos adicionales agregados por las organizaciones que

usan la plataforma. Aunque en uno de los casos analizados se encuentra el campo *género*, este dato no es recuperado por el informe de registro de actividad.

Retomando a Meza Fernández y Marquet (2017), se señala que las variables de género y procedencia disciplinar serían datos significativos para la interpretación. El género no se encuentra por defecto entre los campos de información de usuario/a de *Moodle*. Sin embargo, puede añadirse como campo personalizado o recuperarse desde otra base de datos, por ejemplo, un sistema académico.

Para cualquier pregunta de investigación vinculada a la problemática de brecha digital de género (Castaño, 2008), no sería posible recuperar correlaciones o patrones desagregados, puesto que, o bien el dato sobre género no existe, o no será recuperado por los informes de *Moodle*. Este deberá obtenerse de otra fuente o con otro procedimiento (como reportes configurables de Moodle, que requieren programación y privilegios de usuario), impidiendo la lectura de este tipo de relaciones al usuario con perfil docente.

A modo de síntesis, en el nivel denotativo encontramos que los datos seleccionados por el informe son insuficientes para realizar una valoración significativa de la actividad de los usuarios. En el nivel connotativo observamos que los informes a menudo *infieren* ciertos comportamientos de parte de los estudiantes, los cuales sólo están parcialmente representados por tales registros. Por último, el nivel deconstructivo nos permite visualizar qué otros datos podrían incorporarse para enriquecer una lectura pedagógica y de investigación.

3. Conclusiones

Las tendencias a la digitalización y plataformización se hacen presentes en casi todos los dominios de la vida social, profundizándose con la pandemia global de COVID-19. Junto con ello, el registro de interacciones y la concentración de datos se vuelve un tema central.

El artículo presentó exploraciones en torno al uso de informes, registros y bases de datos provenientes del LMS Moodle. El abordaje se enmarca en un campo emergente de métodos digitales, buscando analizar las potencialidades y límites de estos nuevos recursos. Observamos un creciente interés por la utilización de estas fuentes de datos, con los objetivos de describir y generar predicciones sobre el comportamientos de estudiantes, así como desarrollar herramientas para optimizar el rendimiento y acompañamiento.

Los casos locales relevados muestran que, aunque *Moodle* se expande en diferentes modalidades de enseñanza, sus posibilidades en cuanto recurso de investigación son aún incipientes. Encontramos limitaciones en el acceso a los datos, en tanto la falta de previsión y definiciones sobre su resguardo, tratamiento y uso derivan en que los mismos no reúnan condiciones necesarias para ser consideradas como fuente. Los hallazgos de este trabajo indican algunos lineamientos y buenas prácticas en este sentido.

Junto con ello, buscamos describir los tipos de recursos potencialmente disponibles. Reconocimos aquí dos niveles de análisis: el nivel de cursos y el nivel de usuarios. Mientras el primero permite mapear los recursos utilizados por docentes en sus prácticas de enseñanza, el segundo permite principalmente analizar las interacciones a través de LMS.

El análisis de un registro de actividad de curso señala cierta limitación intrínseca en los datos recogidos por la plataforma para interpretar procesos de aprendizaje. Algunas de esas limitaciones podrían superarse agregando nuevos datos que no se registran o no se recuperan en las consultas predeterminadas. Una indagación cualitativa en torno a esa misma experiencia podría complementar y ofrecer información valiosa sobre los modos de uso y apropiación por parte de diferentes usuarios. Estudios comparativos entre datos normalizados de varios LMS con modalidades de enseñanza similares permitirían, asimismo, contribuir a consolidar patrones de uso y comportamientos frecuentes. Estos son fuente de información valiosa tanto para los docentes como para los equipos técnicos que hacen evaluación de usabilidad de la plataforma.

El ensayo buscó sistematizar el propio acercamiento a los LMS con fines de investigación. Frente a un cierto clima de época donde pareciera que los *datos* están allí, listos para extraer nuevas verdades y conocimientos, es necesario mantener una perspectiva crítica sobre sus alcances y significados. Ello observamos en nuestro recorrido, donde la posibilidad de acceder a un registro pormenorizado de los procesos de enseñanza y aprendizaje se iba desvaneciendo a cada paso. En este sentido, los hallazgos remiten a los casos analizados. Los mismos podrían ampliarse en función de nuevas indagaciones que profundicen en nuevos contextos, plataformas o niveles educativos.

El debate en torno a las analíticas de aprendizaje y su carácter predictivo, útiles para el diseño de cursos o para recomendar acciones a los/as estudiantes, se encuentra en una etapa clave: ¿pueden los datos almacenados en LMS predecir el éxito o fracaso en experiencias de aprendizaje o realizar acciones de seguimiento de estudiantes en forma inteligente? En cualquier caso, para comprender las posibilidades y límites de estos abordajes, resulta necesario avanzar en los desafíos de su acceso e interpretación. El potencial de los datos generados por las plataformas virtuales de aprendizaje amerita las apuestas por este campo interdisciplinar de estudios.

Referencias

- Amo, D., Alier, M., García-Peñalvo, F. J., Fonseca, D., & Casañ, M. J. (2020). Protected users: A moodle plugin to improve confidentiality and privacy support through user aliases. *Sustainability*, 12(6), 2548. <https://doi.org/10.3390/su12062548>
- Ardévol, E. (2016). Big data y descripción densa. *Virtualis*, 7(14), 14-38. <https://doi.org/10.2123/virtualis.v7i14.186>
- Bruns, A. & Burgess, J. (2016). Methodological innovation in precarious spaces: the case of twitter. En H. Snee, C. Hine, Y. Morey, S. Roberts y H. Watson (2016), *Digital Methods for Social Science*. Palgrave Macmillan.
- Campo, M., Amandi, A., & Biset, J. C. (2021). A software architecture perspective about Moodle flexibility for supporting empirical research of teaching theories. *Education and Information Technologies*, 26(1), 817-842. <https://doi.org/10.1007/s10639-020-10291-4>
- Castaño, C. (2008). *La segunda brecha digital y las mujeres jóvenes*. Cátedra.
- CEPAL. (2020). Gestión de datos de investigación. Biblioguías. <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion>
- Digital Methods Initiative (DMI). (2015). Post-Snowden Media Empiricism and Secondary Social Media: Data Studies Beyond Facebook and Twitter. *DMI Summer School 2015*. <https://wiki.digitalmethods.net/Dmi/SummerSchool2015>
- Ferrero, F. (2018). Big data y educación: un análisis vigotskiano de los “algoritmos predictivos del éxito de los estudiantes”. En P. Mercado, N. Gonzalez, L. Beltramino y F. Ferrero (Eds.), *Contextos de producción de la teoría de Lev Vigotsky, a 120 años de su nacimiento*. FFyH-UNC.
- Zanotti, A., & Pagola, L. I. (2020). Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle. *Transdigital*, 3(6), 1–23. <https://doi.org/10.56162/transdigital145>

- Franzke, A. S., Bechmann, A., Zimmer, M., & Ess, C. (2020). *Internet Research: Ethical Guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Gamage, S. H., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education*, 9(1), 1-24. <https://doi.org/10.1186/s40594-021-00323-x>
- Geoidep. (2017). Qué son los metadatos. <https://www.geoidep.gob.pe/metadatos/que-son-los-metadatos>
- Examulator. (2022). *moodle_3.11 Database*. <https://www.examulator.com/er/output/index.html>
- Jenaro-Río, C., Castaño-Calle, R., Martín-Pastor, M., & Flores-Robaina, N. (2018). Rendimiento académico en educación superior y su asociación con la participación activa en la plataforma Moodle. *Estudios sobre educación*, 34(1), 177-198. <https://doi.org/10.15581/004.34.177-198>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1), 1-12. <https://doi.org/10.1177/2053951714528481>
- Kuo, R., Krahn, T., & Chang, M. (2021). Behaviour Analytics-A Moodle Plug-in to Visualize Students' Learning Patterns. En A.I. Cristea, C. Troussas (Eds.), *Intelligent Tutoring Systems*. Springer.
- Lomborg, S. (2012). Researching communicative practice: web archiving in qualitative media research. *Journal of technology in human service*, 30(1), 219-231. <https://doi.org/10.1080/15228835.2012.744719>
- López-Zambrano, J., Lara, J. A., & Romero, C. (2020). Towards portability of models for predicting students' final performance in university courses starting from moodle logs. *Applied Sciences*, 10(1), 354. <https://doi.org/10.3390/app10010354>
- Zanotti, A., & Pagola, L. I. (2020). Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle. *Transdigital*, 3(6), 1-23. <https://doi.org/10.56162/transdigital145>

- Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the digital humanities*, 2(1), 460-475. <https://doi.org/10.5749/minnesota/9780816677948.003.0047>
- Meza Fernández, S., & Marquet, P. (2017). Styles de navigation sur un site Web chilien pour des enseignants d'espagnol. *Canadian Journal of Learning and Technology*, 43(2). <https://cjlt.ca/index.php/cjlt/article/download/27594/20245>
- Moodle. (2022a). About Us. <https://moodle.com/about/>
- Moodle. (2022b). Plugins. <https://moodle.org/plugins/>
- Moodle Docs. (2022a). ad-hoc contributed reports. https://docs.moodle.org/311/en/ad-hoc_contributed_reports
- Moodle Docs. (2022b). Learning Analytics Specification. https://docs.moodle.org/dev/Learning_Analytics_Specification
- Moodle Users Association. (2022). The Committee. <https://moodleassociation.org/committee>
- Open Knowledge International. (s.f.). ¿Qué son los datos abiertos? <http://opendatahandbook.org/guide/es/what-is-open-data/>
- Poell, T., Nieborg, D., & Van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1425>
- Poirier, L. (2021). Reading datasets: Strategies for interpreting the politics of data signification. *Big Data & Society*, 8(2), 1-19. <https://doi.org/10.1177/20539517211029322>
- Rogers, R. (2013). *Digital methods*. MIT press.
- Zanotti, A., & Pagola, L. I. (2020). Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle. *Transdigital*, 3(6), 1–23. <https://doi.org/10.56162/transdigital145>

- Selwyn, N. (2016). *Is technology good for education?* Polity Press.
- Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (2016). *Digital Methods for Social Science*. Palgrave Macmillan.
- Srnicek, N. (2018). *Capitalismo de plataformas*. Caja Negra.
- Sued, G. (2019). Para una traducción de los métodos digitales a los estudios latinoamericanos de la comunicación. *Virtualis*, 10(19), 20-41. <https://doi.org/10.2123/virtualis.v10i19.295>
- Thompson, G. (2017). Computer adaptive testing, big data and algorithmic approaches to education. *British journal of sociology of education*, 38(6), 827-840. <https://doi.org/10.1080/01425692.2016.1158640>
- UNESCO (2021). Proyecto de Recomendación de la UNESCO sobre la Ciencia Abierta. https://unesdoc.unesco.org/ark:/48223/pf0000376893_spa.locale=en
- Van Dijck, J. (2016). *La cultura de la conectividad: una historia crítica de las redes sociales*. Siglo XXI.
- Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
- Vélez, J. G. & Zanotti, A. (2020). La producción en comunidades FLOSS: empresas, Fundaciones y governance. En H. Morero y J. Motta (Eds.), *La economía del software libre y open source: Multinacionales, Pymes y Comunidades*. Estudios Sociológicos.
- Zukerfeld, M. (2020). Bits, plataformas y autómatas. Las tendencias del trabajo en el capitalismo informacional. *Revista Latinoamericana de Antropología del Trabajo*, 4(7), 1-50. <http://hdl.handle.net/11336/168942>
- Zanotti, A., & Pagola, L. I. (2020). Potencialidades y límites para el análisis de datos de sistemas de gestión de aprendizaje. El caso de Moodle. *Transdigital*, 3(6), 1-23. <https://doi.org/10.56162/transdigital145>