Learning Analytics y protección de datos personales. Recomendaciones.

Patricia Díaz [1], **Matías Jackson** [2] **Regina Motz** [3]

[1] Docente de Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay

[2] Docente del Instituto de Derecho Informático, Facultad de Derecho, Universidad de la República, Uruguay

[3] Docente de Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay

pdiaz@fing.edu.uy, matias@mjackson.uy, rmotz@fing.edu.uy

Abstract. *Data analytics techniques provide excellent tools to understand student learning patterns and to improve the instructional design, organizing the content and activities hosted in virtual platforms in a more efficient way. But we must not forget that most of the processed data contain personal digital information related to students behavior. According to the LATAM countries regulatory framework, there are limitations and principles that must be respected when we collect, process, store and analyze personal data. In this article we summarize the basic legal recommendations for automated processing of personal data in learning analytics related activities.*

Resumen. *Las técnicas de análisis de datos automatizadas proveen excelentes herramientas para comprender los patrones aprendizaje de los estudiantes y mejorar el diseño instruccional organizando de forma más eficiente los contenidos y actividades de los cursos alojados en plataformas virtuales. Pero no debemos olvidar que la mayoría de los datos procesados contienen información personal relacionada con el comportamiento de nuestros estudiantes. De acuerdo con el marco regulatorio de la mayoría de los países de LATAM existen limitaciones y principios a cumplir al momento de recolectar, procesar, almacenar y analizar datos personales. En este artículo intentaremos resumir las recomendaciones legales básicas para el tratamiento de automatizado de datos personales en el marco de actividades de learning analytics.*

DOI: 10.5753/cbie.wcbie.2015.981

1. Introducción

El actual progreso tecnológico nos permite capturar todos los eventos que suceden en los ambientes de aprendizaje, cada interacción o ingreso a un recurso pueden ser recogidos y almacenados. Como consecuencia de ello estos ecosistemas de aprendizaje hoy pueden ser analizados utilizando técnicas de minería de datos. Estas técnicas de análisis de datos automatizadas constituyen una excelente herramienta para comprender los patrones de aprendizaje de los estudiantes y mejorar el diseño instruccional organizando de forma más eficiente los contenidos y actividades de los cursos alojados en plataformas virtuales. Pero al momento de utilizarlas no debemos olvidar que la mayoría de los datos procesados contienen información personal relacionada con el comportamiento de nuestros estudiantes y que tenemos la obligación de preocuparnos por preservar su privacidad.

En la mayoría de los países de Latinoamérica éste tipo de información personal se regula a través de leyes de Protección de Datos Personales. Brasil, Colombia, Paraguay, Perú, Argentina, Ecuador, Panamá y Honduras han reconocido el Habeas Data como derecho constitucional. Argentina, Uruguay, México, Perú, Costa Rica y Colombia han promulgado leyes de protección de datos basadas en la Directiva de la UE de 1995[1]. Chile y Paraguay cuentan con leyes de protección de datos aunque no cuentan con autoridad de protección de datos. En base a estas leyes encontramos que existen limitaciones y principios a cumplir al momento de recolectar, procesar, almacenar y analizar los datos personales.

En este artículo brindaremos recomendaciones básicas relacionadas con los aspectos jurídicos del tratamiento automatizado de datos personales provenientes plataformas de aprendizaje.

2. ¿Qué es un dato personal?

Un dato personal es **cualquier tipo de información que pueda identificar directamente o indirectamente a una persona**, ya sea un nombre, dirección, teléfono, cédula de identidad, número de estudiante, la fotografía de un estudiante o una muestra caligráfica.

Cuando las bases de datos contienen información relacionada con una persona identificada o identificable son reguladas por las leyes de Protección de Datos Personales.

Las técnicas de minería de datos en ambientes de *e-learning* permiten crear un seguimiento de cada estudiante recogiendo sus participaciones, hábitos de uso, cantidad

[1] DIRECTIVA 95 / 46 / CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 24 de octubre de 1995 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Disponible en: http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:31995L0046&from=en

de horas online, frecuencia de tipeo en el teclado, etc. Mediante la recolección de estas características se pueden generar perfiles que permiten identificar a cada estudiante, y que luego son almacenados en bases.

Estos datos refieren a personas físicas, y por tanto se encuentran protegidos por la normativa general de protección de datos personales.

3. ¿Qué principios generales los rigen?

En la mayoría de los países latinoamericanos, la protección de datos personales implica la existencia de los siguientes principios[2]:

1. **Previo consentimiento informado**: se trata de uno de los principios rectores sobre el manejo de datos personales que todos deberíamos conocer.

   Los datos no pueden ser recabados, procesados, publicados o comunicados sin contar con el consentimiento del usuario. Este consentimiento debe ser:

   a. Libre (podrá brindarse o no).

   b. Previo (recabado antes de solicitar los datos).

   c. Expreso (no tácito o implícito).

   d. Documentado (verificable).

   e. Informado (deberá conocerse la finalidad por la que se recolectan los datos y dónde se podrán ejercer los derechos de control).

   Comúnmente constituye una excepción al principio el consentimiento informado el hecho de que la información se recolecte para el uso exclusivo, personal, individual o doméstico. Es por esto que, cuando el análisis de datos se efectúa por el propio docente con el único fin de mejorar su desempeño docente y los datos no se publican, comparten o conservan, no será necesario solicitar autorización. Si, por el contrario, el análisis de datos se incorpora como una actividad de mejora continua por una Institución Educativa (IE) o con fines de investigación, será necesario contar con el consentimiento informado de los implicados.

[2] Los principios enumerados en el presente artículo son una adaptación al contexto de los países de LATAM de los previstos en la "*Propuesta Conjunta para la Redacción de Estándares Internacionales para la protección de la Privacidad, en relación con el Tratamiento de Datos de carácter personal*. Resolución de Madrid, 2009."

2. **Legalidad.** Las bases de datos personales deben cumplir con la normativa y en algunos países deben inscribirse en un registro.

3. **Veracidad.** Los datos registrados deberán ser veraces, adecuados, ecuánimes (imparciales) y no excesivos en relación con la finalidad para la que se han obtenido.

4. **Finalidad.** Los datos no deben utilizarse para fines diferentes a los solicitados. Cumplida su finalidad, deben eliminarse.

5. **Seguridad.** La normativa comúnmente señala qué medidas mínimas de seguridad se deben adoptar para proteger los datos recolectados.

6. **Reserva.** Los datos deben utilizarse únicamente para la finalidad con la que se obtuvieron, y aplica el deber de confidencialidad a personas que tengan acceso a los mismos.

7. **Responsabilidad.** Recae sobre la persona física o jurídica responsable de la base así como los encargados de tratamientos, usuarios y terceros, con diferente alcance.

4. ¿Qué derechos tienen los usuarios de datos personales?

El usuario de datos tiene el control sobre estos. Este control incluye el derecho a acceder a sus datos, rectificarlos o actualizarlos, solicitar su inclusión en una base, solicitar la supresión de los datos que le perjudiquen u oponerse al tratamiento de estos.

Tal como expresa Mejan (1994), debemos tomar en cuenta que el riesgo de frialdad, injusticia y despersonalización aumenta cuando la información no se queda en el terreno documental sino que se traslada al decisional. Es por ello que cualquier estudiante podrá impugnar aquellas valoraciones personales, que lo afecten de manera significativa y que se basen únicamente en un tratamiento automatizado de datos que evalúan determinados aspectos como el rendimiento académico o su conducta.

5. Recomendaciones

El objetivo de este artículo será presentar una serie de recomendaciones básicas, dirigidas a equipos de investigación que trabajen en el contexto específico de learning analytics, a los efectos de incorporar los principios relacionados la protección de datos

a sus actividades. También propondremos recomendaciones básicas dirigidas a IE que decidan incorporar políticas de learning analytics.

5.1. Consentimiento informado y Reserva

Si decidimos utilizar técnicas de learning analytics y nuestra IE no cuenta con una política estandarizada que regule el tema, será conveniente la redacción de un aviso a los estudiantes a los efectos de obtener el consentimiento informado, especificando el uso que se le dará a los datos que resulten del o los curso/s y que estos serán anonimizados. Por ejemplo:

> *Aviso: Los datos como estadísticas, cantidad de clics, enlaces visitados, materiales descargados, calificaciones, participaciones en foros y, en general, el historial del recorrido de cada estudiante en este curso serán procesadas de forma automática en un sistema que utiliza esos datos para recomendar cambios y mejoras en el funcionamiento y organización de ésta plataforma de aprendizaje. El responsable del almacenamiento y procesamiento de estos datos será "Xxxxxx" y su mail es: xxxx@xxxx.xxx.*
>
> *Se aplicarán técnicas de anonimización de sus datos en todas las publicaciones.*

De acuerdo con el principio del "previo consentimiento informado" recomendamos la utilización de un mecanismo de recolección automatizada de consentimiento, como por ejemplo una encuesta obligatoria al inicio del curso. De esta forma estaremos solicitando que el estudiante brinde su consentimiento expreso y documentado. Vale la pena resaltar que si, como resultado del procesamiento de los datos, se deriva alguna valoración personal, que afecte de manera significativa al estudiante, este tiene derecho a ser informado sobre el criterio de valoración y el programa utilizado para ello.

Si la información recolectada (la base original en "crudo") será compartida con otro equipo de investigadores o con otra institución y esto no fue informado a los estudiantes, se deberá solicitar nueva autorización explicitando con quién/quiénes se compartirá, para qué fines se volverá a utilizar, cómo se procesará, dónde y por cuánto tiempo será almacenada por esos terceros y cómo contactar al responsable del tratamiento.

Algunos ordenamientos además prevén que, si estas instituciones con las que se piensa compartir la información se encuentran en el exterior, se debe garantizar una legislación adecuada en materia de protección de datos en el país de destino. En ese caso debemos consultar a la Autoridad de Protección de Datos Nacional cuáles son los países con los que podremos compartir datos.

Cuando se trata de datos de menores de edad, el consentimiento deben otorgarlo los padres o tutores.

5.2. Reserva, Responsabilidad y Seguridad

Anonimización de datos - Cada vez que se compartan los datos o los resultados del análisis de datos con terceros, deberán anonimizarse (de esta forma no será necesario volver a recabar el consentimiento informado de los titulares de datos).

En lo posible, se deberán utilizar métodos de anonimización que no sean fáciles de vulnerar mediante técnicas de triangulación de datos. No es correcto afirmar que "se garantizará la total anonimización de los datos", porque cada vez más investigaciones demuestran que:

- este objetivo es muy difícil de lograr (Narayanan y Shmatikov, 2010)
- los datos pueden ser perfectamente anónimos o perfectamente útiles, casi nunca ambos a la vez; (Ohm, 2010).

Durante el proceso de anonimización de datos se recomienda eliminar las direcciones IP ya que este es un dato de fácil triangulación y existen antecedentes en los que se considera la dirección IP como un dato personal[3], dada la posibilidad de constituir un identificador único en terminales de telecomunicaciones.

Preservación del dato: el responsable de la base de datos deberá restringir el acceso administrativo a las bases que contengan datos personales.

5.3. Finalidad y Conservación de los datos

Como principio general, no se deben conservar los datos sin anonimizar (la base "cruda") si no es necesario.

El principio de finalidad nos indica que, una vez que se extingue la razón original por la que fueron recolectados los datos, estos deben eliminarse. Por lo que las instituciones educativas que optan por aplicar learning analytics de forma sistemática deberán explicitar su política de conservación de datos, decidiendo, por ejemplo, si existen razones que justifiquen mantener en sus bases la información de los egresados.

Como dijimos anteriormente, en muchos países existe la obligación de Registro de las Bases de Datos para que estas sean consideradas "legales". En estos casos, si se decide conservar los datos, la base deberá ser debidamente registrada.

5.4. Datos sensibles

Debemos tomar en cuenta que la mayoría de las legislaciones prohíbe el procesamiento de datos personales que revelen origen racial y étnico, preferencias políticas, convicciones religiosas o morales, afiliación sindical e informaciones referentes a la salud o a la vida sexual sin un permiso específico. Por lo que, si recolectamos este tipo de información, debemos tener cuidado de que permanezca anónima en nuestra base o

[3] Artículo 29: Grupo de Protección de Datos de la UE. 10750/02/ES/FinalWP 58. Dictamen 2/2002 sobre el uso de identificadores únicos en los equipos terminales de telecomunicaciones: ejemplo del IPv6. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2002/wp58_es.pdf

solicitar el permiso correspondiente, registrando nuestra base de acuerdo a lo que disponga la legislación de nuestro país.

5.5. Instituciones Educativas

Cuando se incorporan políticas de learning analytics en IE se sugiere:

- la creación de un comité de ética que se encargue de: 1) implementar guías con buenas prácticas en el manejo de datos personales para los docentes que adopten técnicas de learning analytics y 2) analizar los posibles conflictos que vayan surgiendo.
- la publicación, de la forma más clara posible, de la política de manejo de datos que aceptan los estudiantes al formar parte de la IE. Esta deberá incluir al menos siguientes aspectos: ¿qué datos se recogen?¿por cuánto tiempo se conservarán?, ¿con qué fines se utilizarán?, ¿qué política de seguridad adopta la institución?, ¿quiénes son los responsables del tratamiento de los datos?
- fomentar la transparencia mediante la incorporación de un procedimiento adecuado para que el estudiante pueda acceder a sus datos y solicitar la rectificación (si existen errores).
- restringir el acceso administrativo a las bases que contengan datos personales.
- elevar consulta, ante cualquier duda, a la autoridad nacional de protección de datos (atención: esta autoridad no existe en todos los países de LATAM).

Plataformas diseñadas y gestionadas por la propia IE: resulta importante que las IE tomen en cuenta las cuestiones de privacidad de datos utilizando modelos conocidos como “privacy by design” en los que estas cuestiones se incluyen como uno de los requerimientos iniciales al pensar el diseño de un sistema (Le Métayer, 2010). Un ejemplo de la aplicación de esta filosofía sería la implementación de aplicaciones de manejo de datos que diferencien datos personales de datos NO personales al momento de diseñar las herramientas en un modelo de análisis.

Plataformas diseñadas y gestionadas por terceros: es una tendencia cada vez mayor la contratación de servicios de nube por parte de las IE. Estos servicios pueden cubrir tanto la de gestión administrativa (almacenamiento de datos y cuentas de correo) como las plataformas de aprendizaje (servicios de aulas virtuales y aplicaciones educativas) y, muchas veces, incluyen algún tipo de servicio de análisis de datos. Destacamos que las IE son las responsables finales de los datos de los docentes y estudiantes por lo que, al momento de seleccionar la empresa que brindará el servicio, deberán analizar al menos: a) la Política de privacidad y antecedentes de dicha empresa, b) en qué se basa el modelo de negocio de la empresa, c) dónde se encuentran ubicados los datos y d) cuál será la legislación y jurisdicción aplicable al contrato y qué garantías ofrece. En los países de Latinoamérica aún no son comunes los relevamientos o inspecciones de carácter preventivo en relación al uso de servicios cloud en educación por parte de las autoridades de protección de datos locales. Es por eso que, sugerimos revisar las

recomendaciones propuestas por la Agencia Española de Protección de Datos en su *Informe de resultados de la primera inspección sectorial sobre cuestiones de privacidad en Europa sobre servicios cloud en el ámbito educativo* (julio de 2015).

6. Conclusiones

La necesidad de conciliar los avances tecnológicos en entornos educativos con los derechos y libertades fundamentales de los diferentes actores involucrados es un tema que no admite postergación; para ello, se aconseja la adopción de una actitud proactiva a favor de la protección de datos personales por parte de las instituciones educativas e investigadores en el marco del desarrollo de actividades de learning analytics.

Finalmente, consideramos que las recomendaciones desarrolladas deberían tomarse en cuenta independientemente de que la legislación del país en que se emprenderán actividades de learning analytics prevea un régimen de protección de datos, pues el respeto por la privacidad es también una cuestión de ética profesional.

6. Referencias

Agencia Española de Protección de Datos (2015) en su "Informe de resultados de la primera inspección sectorial sobre cuestiones de privacidad en Europa sobre servicios cloud en el ámbito educativo". Disponible en: https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/Inspeccion_cloud_educacion.pdf

Conferencia Internacional de Autoridades de Protección de Datos y Privacidad (2009) "Propuesta Conjunta para la Redacción de Estándares Internacionales para la protección de la Privacidad, en relación con el Tratamiento de Datos de carácter personal." Resolución de Madrid, 2009.. Disponible en: https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/

Documents/Cooperation/Conference_int/09-11-05_Madrid_Int_standards_ES.pdf,

julio 2015.

Grupo del Artículo 29 _Grupo de Protección de Datos de la UE. (2002) "Dictamen 2/2002 sobre el uso de identificadores únicos en los equipos terminales de telecomunicaciones: ejemplo del IPv6." 10750/02/ES/FinalWP 58. Disponible en: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2002/wp58_es.pdf, julio de 2015.

Le Métayer, Daniel (2010), "Privacy by Design: a Formal Framework for the Analysis of Architectural Choices (extended version)", Research Report, RR-8229, 2013, pp.24. <hal-00788584>. Disponible en: https://hal.archives-ouvertes.fr/hal-00788584/document, julio 2015.

Narayanan, Arvind y Shmatikov, Vitaly (2010), "Privacy and security Myths and fallacies of Personally identifiable information", Communications of the ACM, june 2010 vol. 53 no. 6. Disponible en: https://www.cs.utexas.edu/~shmat/shmat_cacm10.pdf, Julio 2015.

Meján, Luis Manuel C. (1994), El derecho la intimidad y la informática, México, Porrúa.

Ohm, Paul, Broken Promises of Privacy (2009), "Responding to the Surprising Failure of Anonymization", en UCLA Law Review, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. Disponible en SSRN: http://ssrn.com/abstract=1450006, junio 2015.

Pardo, Abelardo y Siemens, George (2014), "Ethical and privacy principles for learning analytics", en British Journal of Educational Technology, Vol 45 No 3 2014, pags 438–450. Disponible en: http://www.researchgate.net/publication/261331318_Ethical_and_privacy_principles_for_learning_analytics, junio 2015.