Learning Analytics et recherche scientifique au prisme du droit des données personnelles : contribution d'un projet de recherche en Éducation aux Médias et à l'Information

Tanja Petelin[1], Hassina El Kechaï[2]

[1] Université de Poitiers, CECOJI, UR- 21665 Poitiers, France
tanja.petelin@ univ-poitiers.fr
[2] Université de Poitiers, TECHNE, UR- 20297 Poitiers, France
hassina.el.kechai@ univ-poitiers.fr

Résumé. Le travail décrit dans cet article s'inscrit dans le contexte des *Learning Analytics* et porte sur la mise en œuvre des orientations majeures du RGPD dans le cadre des recherches en EIAH ayant recours à la collecte de données à caractère personnel. Il décrit un cas d'usage dans le contexte de la conception d'un dispositif dans le domaine de l'EMI (Éducation aux Médias et à l'Information) en contexte scolaire pour un public de lycéens. Ce dispositif porte sur la collecte de leurs données d'interaction issues de l'utilisation de leurs Smartphones et leur restitue leurs pratiques numériques sous forme de tableaux de bord. Nous nous focalisons dans cet article sur la mise en conformité du RGPD du dispositif que nous illustrons par quelques exemples.

Mots-clés : RGPD, donnée à caractère personnel, Éducation aux Médias et à l'Information, *Learning Analytics*

Abstract. The article provides a proposal of guidelines describing the main principles of GDPR compliance in the context of EIAH research that involves the collection of personal data. The applicative context of this guide is in the design of a device in the field of EMI (Media and Information Education) in a school context for a high school audience and deals with the collection of their interaction data from smartphones. We provide some examples of the implementation of GDPR in this context.

Keywords: GDPR, personal data, Media and Information Education, Learning Analytics.

1 Introduction

Le Règlement Général sur la Protection des Données à caractère personnel "RGPD" [12] est entré en vigueur le 25 mai 2018. Ce texte, d'application directe dans tous les États membres de l'Union européenne, laisse néanmoins à ces derniers une marge de

manœuvre sur certains points. En France, la loi informatique et libertés est ainsi maintenue : en complément du règlement, elle intègre notamment des dispositions d'adaptation du RGPD au contexte national [10]. Le RGPD concerne tout acteur privé ou public établi sur le territoire de l'Union européenne, que le traitement ait lieu ou non dans l'Union, ainsi que ceux non établis dans l'Union dès lors que les activités de traitement ciblent des personnes concernées qui se trouvent sur le territoire de l'Union (RGPD, art. 3). Les travaux de recherche en EIAH ont de plus en plus recours à la collecte et à l'analyse de données d'apprentissage. Ces travaux ont rendu populaire le champ disciplinaire des *Learning Analytics*, ou « analyse de l'apprentissage » qui vise à recueillir les données liées à un dispositif d'apprentissage, à les analyser puis à présenter les résultats de l'analyse sous forme de rapports par exemple ou de tableaux de bord [5]. Pour notre part, nous collectons les données pour réaliser des analyses d'usages et la restitution de ces analyses peut s'adresser à des apprenants pour les mettre dans une démarche réflexive sur leurs pratiques ou à des enseignants pour réguler des pratiques numériques. Cela peut également servir à des institutions pour mesurer la pertinence d'une politique d'équipement numérique, pour produire des connaissances sur les pratiques numériques des élèves ou encore pour accompagner les transformations pédagogiques des enseignants [11]. Nous nous inscrivons donc dans l'analyse des comportements en termes d'utilisation du numérique et en aucun cas sur l'analyse de la performance en termes d'apprentissage. Tous ces travaux sont tenus de s'assurer que le traitement des données personnelles collectées au travers de différents dispositifs de *Learning Analytics* soit conforme au RGPD. Traiter les données conformément au RGPD signifie qu'il faut encadrer les données tout le long de leur cycle de vie à savoir : leur collecte, leur conservation, leur exploitation, leur modification, leur communication, leur archivage, jusqu'à leur destruction ou anonymisation. C'est ainsi que le RGPD doit non seulement accompagner tout le processus de conception et de développement de dispositifs utilisant les *Learning Analytics*, mais peut aller jusqu'à les façonner en termes d'interfaces notamment.

Alors que le RGPD est parfois perçu comme une contrainte[1], pouvant limiter la mise en œuvre des objectifs de la recherche, son analyse plus approfondie révèle le contraire. En effet, le RGPD vise à accorder aux personnes physiques un contrôle sur leurs données personnelles et impose aux responsables de traitement une démarche de conformité, qui doit être rigoureusement respectée et documentée (RGPD, art. 24). Pourtant, loin de constituer un obstacle, chacune des étapes offre une certaine souplesse. Les besoins particuliers de la recherche scientifique donnent d'ailleurs lieu à quelques dérogations[2], tout en rappelant, dans ce cadre, l'obligation de mettre en place des garanties appropriées pour les droits et libertés des personnes (RGPD, art. 89). Le RGPD constitue ainsi un cadre obligatoire, basé sur un socle de principes directeurs, la responsabilisation des acteurs et le respect des droits des personnes concernées : sa mise en œuvre concrète dans le domaine des EIAH peut constituer une opportunité pour le développement d'une recherche respectueuse des droits des personnes sans sacrifier les objectifs scientifiques.

[1] Complexité du travail avec les DPO, difficulté de trouver des solutions ciblées et souples dans les articles du RGPD, changement de posture du public cible dû à la connaissance des finalités de traitement de la recherche imposée par le RGPD sur le principe de transparence...

[2] V. notamment art. 5.1 b) (limitation des finalités) et e) (limitation de la durée de conservation).

Dans cet article, nous traitons de la problématique de la prise en compte du RGPD dans la conception de dispositifs de Learning Analytics et de sa mise en œuvre dans le domaine des EIAH. Pour répondre à cette problématique et après la présentation d'un état de l'art qui traite de la protection des données personnelles et des règles de conduite dictées par le RGPD, nous présentons les orientations importantes à prendre en compte pour s'assurer que les dispositifs soient conformes aux règles établies par le RGPD. Nous proposons ensuite quelques exemples de mise en œuvre de ces orientations dans le cadre du processus de conception et de développement d'un dispositif en Éducation aux Médias et à l'Information (EMI) dans le contexte scolaire.

2 État de l'art

2.1 Protection des données personnelles

Le RGPD définit une donnée personnelle comme « toute information se rapportant à une personne physique identifiée ou identifiable » (art. 4.1). Ces informations peuvent être des identifiants, tels qu'un nom, un numéro d'identification, les données de localisation, un identifiant en ligne, ou encore des informations plus spécifiques, propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. Toutes les données personnelles ne sont pas à mettre au même niveau. En effet, certaines peuvent être considérées comme "sensibles" et peuvent être soumises à un régime juridique de protection renforcée, basée sur le principe d'interdiction de traitement, assorti d'une liste des exceptions. Les données dites sensibles peuvent être par exemple : l'origine raciale ou ethnique, l'opinion politique ou religieuse, l'appartenance syndicale, les données relatives à la santé, à la vie ou à l'orientation sexuelle (RGPD, art. 9).

Le RGPD s'applique dès lors qu'un dispositif traite des données personnelles, ce dernier terme étant défini comme « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés [...] telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction » (RGPD, art. 4.2).

Le traitement des données personnelles doit être distingué du traitement des données anonymes ou anonymisées : ces dernières ne rentrent pas dans le champ d'application de la protection des données personnelles. La technique d'anonymisation doit cependant correspondre aux exigences strictes de la CNIL et « empêcher toutes les parties d'isoler un individu dans un ensemble de données, de relier entre eux deux enregistrements dans un ensemble de données (ou dans deux ensembles de données séparés) et de déduire des informations de cet ensemble de données » [4] [7]. L'anonymisation est ainsi incompatible avec les traitements dont la finalité est le suivi du comportement ou de l'activité d'une personne [4]. En effet, cette technique ne doit pas être confondue avec la pseudonymisation, consistant à remplacer le traitement des informations directement identifiables (nom, adresse IP...) par des données non directement identifiables (p. ex. un numéro aléatoire). Dans ce second cas, il est possible de retrouver l'identité

de la personne si l'on dispose d'informations supplémentaires : la législation sur la protection des données personnelles s'applique.

2.2 RGPD et applications de *Learning Analytics* : règles de conduite

Lorsque l'on conçoit ou l'on utilise des dispositifs de *Learning Analytics*, le principe de *Privacy by design* impose la mise en œuvre des principes relatifs à la protection des données tant au moment de la détermination des moyens du traitement qu'au moment du traitement lui-même (RGPD, art. 25.1). Les chercheurs en EIAH qui conçoivent des dispositifs intégrant des *Learning Analytics* doivent donc prendre en compte les principes et règles de protection des données personnelles dès la phase de conception des dispositifs de collecte et d'analyse des données. Selon l'article 5 du RGPD, le traitement des données doit respecter :

- **Les principes de licéité, de loyauté et de transparence :** En premier lieu, ceci impose de disposer d'un *fondement de licéité* qui, en matière de *Learning Analytics*, sera souvent le consentement de la personne concernée. Toutefois, certaines applications de *Learning Analytics* pourraient être basées sur d'autres fondements, cités dans l'article 6 du RGPD, tels que l'intérêt légitime du responsable de traitement ou d'un tiers (à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée) ou, le cas échéant, mission d'intérêt public (une telle mission doit néanmoins être définie par le droit national ou européen – elle ne se présume pas) voire même sur une obligation légale si, dans l'avenir, le législateur impose à certains acteurs de procéder à l'analyse des données d'apprentissage. Quant à la transparence et la loyauté, leur mise en œuvre repose sur une démarche active de communication, compréhensible et aisément accessible, de toute information pertinente, conformément aux exigences du RGPD (art. 12 à 14). Par ailleurs, la personne concernée ne doit jamais être prise au dépourvu à un stade ultérieur quant à la façon dont ses données à caractère personnel ont été utilisées [9].
- **Le principe de limitation des finalités du traitement :** Les données doivent être collectées à des fins préalablement déterminées, explicites et légitimes. Elles ne doivent pas être traitées ultérieurement d'une manière incompatible avec ces finalités. Toutefois, le RGPD prévoit que le traitement ultérieur à des fins de recherche scientifique n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales.
- **Le principe de minimisation des données (ou de nécessité)** : Seules les données adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités peuvent être collectées.
- **Le principe d'exactitude** : Les données doivent être exactes, en cas d'inexactitude, les données doivent être effacées ou rectifiées.
- **Le principe de limitation de la conservation** : Les données ne doivent pas être conservées pour une durée supérieure à ce qui est nécessaire au regard de la finalité du traitement.

- **Les principes d'intégrité et de confidentialité :** Les données doivent être collectées et conservées selon des procédés garantissant la protection contre les intrusions, la perte et la destruction des données.

L'obligation de mise en conformité s'applique en premier lieu au « responsable du traitement » défini comme « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui [...] détermine les finalités et les moyens du traitement ». Dans le cadre des recherches universitaires, les laboratoires ne disposant pas de personnalité juridique, cette qualité revient formellement à l'université dont relèvent les chercheurs. Lorsque les finalités et les moyens du traitement sont définis conjointement par plusieurs responsables de traitement (par exemple dans le cadre d'une collaboration entre plusieurs laboratoires ou avec des structures privées), ceux-ci ont la qualité de responsables conjoints, qui doivent définir de manière transparente leurs obligations respectives au regard du RGPD (art. 26). Enfin, lorsqu'un acteur procède à des traitements des données personnelles en tant que sous-traitant, c'est-à-dire pour le compte d'un responsable de traitement, le RGPD impose des obligations spécifiques (art. 28). Le statut de chacun des intervenants au regard du RGPD dépendra donc de la question de savoir s'il participe ou non à la détermination des finalités et des moyens de traitements.

3 Accompagner l'intégration du RGPD dans la conception de dispositifs de *Learning Analytics* en EIAH

3.1 Démarche de conformité : orientations générales

Malgré le travail d'accompagnement effectué par la CNIL, le RGPD suscite encore de nombreuses interrogations et continue de poser des difficultés dans les projets de conception et d'utilisation de dispositifs de *Learning Analytics* soit par manque de ressources, soit par méconnaissance de la réglementation. Le RGPD a en effet bousculé la manière de concevoir et de développer des dispositifs intégrant des *Learning Analytics*. Bien que des efforts sensibles soient effectués, les chercheurs ont le sentiment qu'il est difficile de s'approcher d'un niveau de conformité élevé sans impacter les objectifs de leurs recherches. Or ce sentiment peut résulter d'un manque de maturité des travaux de recherche en EIAH en termes de mise en conformité : les mécanismes permettant de concilier la protection des données personnelles avec les finalités de la recherche scientifique ne sont pas encore suffisamment mis en œuvre et des incertitudes persistent. Afin d'aider les chercheurs en EIAH à bien intégrer les principes de conformité dans les dispositifs de *Learning Analytics*, nous présentons ci-dessous un guide des principales orientations à prendre en compte et à contextualiser.

Première étape : Dans le cadre d'une recherche menée au sein d'une université ou d'un organisme public, une des premières étapes consiste à se rapprocher du délégué à la protection des données (DPO). Ce dernier est le coordinateur de la mise en conformité au sein de la structure et gère notamment le registre des activités de traitement de

celle-ci, auquel devra être inscrit le traitement envisagé en *Learning Analytics*. Il veillera à ce que le traitement soit conforme au RGPD et sollicitera toute information permettant de démontrer cette conformité, souvent à travers un questionnaire à remplir.

Deuxième étape : Après l'identification des personnes chargées de la réalisation du traitement et la détermination de ce dernier, il faudra expliquer, conformément au *principe de limitation des finalités* de traitement (cf fig.1, point 1), la raison d'être de celui-ci. En effet, comme vu précédemment, les données personnelles ne peuvent être traitées que pour des finalités déterminées, explicites et légitimes. La finalité ainsi définie par les chercheurs en EIAH aura un impact sur la détermination des données qui pourront être traitées, sur la durée de leur conservation, ainsi que sur la définition du fondement de licéité du traitement. Le rôle central de la détermination des finalités et son interaction avec la mise en œuvre des autres principes sont illustrées dans la figure n° 1.

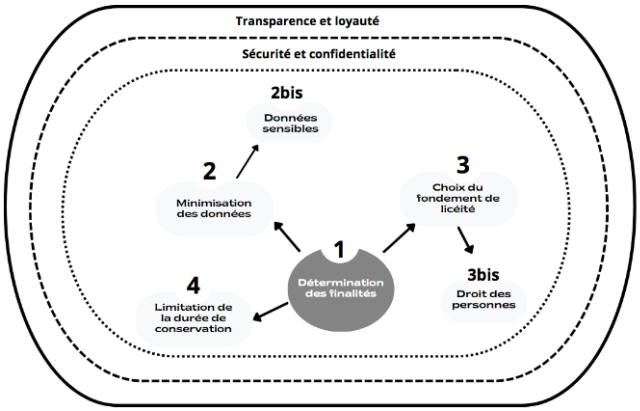


Fig. 1. Les orientations de mise en œuvre des principes du RGPD en EIAH.

Premièrement, selon le *principe de minimisation des données* (cf fig.1, point 2), la finalité justifiera l'étendu des données collectées : avant la mise en œuvre de la collecte, les chercheurs devront ainsi apprécier l'adéquation des données collectées par rapport aux finalités de la recherche et éviter une collecte excessive des données, allant au-delà de ce qui est nécessaire. Par ailleurs, la finalité de la recherche peut parfois imposer la collecte des données sensibles (cf fig.1, point 2bis), telles que définies dans l'article 9 du RGPD et citées plus haut (opinions politiques, religion, santé...). L'identification de ces données n'est pas toujours aisée, car le caractère sensible d'une donnée est parfois indirect. Or, selon un récent arrêt de la Cour de justice de l'Union européenne [2], le régime des données sensibles semble s'y appliquer. Nous pouvons notamment citer les données de géolocalisation qui peuvent par exemple révéler la fréquentation d'un lieu de culte et donc révéler la religion réelle ou supposée d'une personne. Pourtant, ce type de données a parfois un intérêt pour la recherche. Ainsi, dans le projet d'EMI, présenté en section 3.2, les données de géolocalisation permettent d'analyser l'itinérance des pratiques numériques des lycéens et leur circulation dans les sphères personnelle et scolaire, identifiées par les lieux fréquentés. Le traitement des données sensibles étant en principe interdit, une telle conception large des données sensibles peut remettre en

question les pratiques des chercheurs. Toutefois, des exceptions autorisant le traitement de ces données existent (par exemple le consentement explicite de la personne ou, plus spécialement pour la recherche publique, des motifs d'intérêt public important exigeant néanmoins un avis motivé et publié de la CNIL) et peuvent être mobilisées dans le cadre d'une recherche en EIAH. Il convient de rappeler que les conditions de ces exceptions doivent être réunies en plus des conditions du *fondement de licéité* appliqué (V. ci-dessous).

Deuxièmement, la finalité de la collecte aura également un impact sur la détermination du *fondement de licéité* (cf fig.1, point 3). En plus du consentement de la personne concernée, l'article 6.1 du RGPD prévoit cinq autres fondements : si le traitement est nécessaire à la réalisation d'une des fins prévues dans ces cinq hypothèses, ce fondement peut être retenu (nous avons déjà évoqué, en matière de *Learning Analytics*, l'exécution d'une mission d'intérêt public ou la poursuite d'un intérêt légitime) ; dans le cas contraire, le seul fondement rendant licite le traitement est le consentement de la personne concernée (voire de son représentant légal en cas de mineurs). Le choix du fondement de licéité aura un impact sur les droits que le RGPD (art. 15 à 21) reconnaît aux personnes concernées (cf fig.1, point 3bis). Ainsi, lorsque le traitement est fondé sur le consentement, la personne concernée a le droit de le retirer à tout moment, de manière simple et équivalente à celle utilisée pour le recueillir. À partir du retrait du consentement, en principe, tout traitement des données doit cesser et ces dernières doivent être effacées. Pourtant, la suppression des données risque parfois de compromettre les objectifs de la recherche. Cette considération a donné lieu à une exception en matière de recherche scientifique, permettant de déroger à l'obligation d'effacement lorsque ceci peut rendre impossible ou compromettre gravement la réalisation des objectifs du traitement. Alors que le champ d'application concret de cette exception reste à déterminer, lorsque la recherche peut utiliser les données anonymisées, cette solution est à privilégier. Toutefois, l'anonymisation constitue aussi un traitement des données personnelles, qui doit respecter les exigences du RGPD et être basé sur un des fondements de licéité [7]. Le consentement pour le traitement initial ayant été retiré, il serait éventuellement possible, nous semble-t-il, de baser l'anonymisation sur l'intérêt légitime du responsable du traitement (poursuite de la recherche) tout en informant les personnes concernées (transparence). Ce problème, propre aux traitements fondés sur le consentement, ne se pose pas pour les traitements initialement fondés sur la mission d'intérêt public ou sur l'intérêt légitime. Or, bien que l'article 21 du RGPD prévoie un droit d'opposition dans ces deux hypothèses, ce droit n'est pas absolu. En effet, selon cet article, le traitement peut être poursuivi s'il « *existe des motifs légitimes et impérieux pour le traitement qui prévalent sur les intérêts et les droits et libertés de la personne concernée* ». Plus spécifiquement, dans le domaine de la recherche scientifique, il est possible d'y déroger lorsque l'exercice de ce droit risque de rendre impossible ou d'entraver sérieusement la réalisation des finalités de la recherche[3]. Ces deux fondements de licéité peuvent donc offrir plus de sécurité aux chercheurs quant à la possibilité de mener à terme la recherche nécessitant des données personnelles. Toutefois, leurs conditions ne sont pas toujours réunies, la mission d'intérêt public devant être explicite-

[3] RGPD, art. 89.2 ; LIL, art. 78 ; Décr. n° 2019-536 du 29 mai 2019, art. 116

ment prévue par le droit et l'intérêt légitime ne pouvant justifier le traitement que lorsque *« ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel, notamment lorsque la personne concernée est un enfant* » (balance des intérêts).

Enfin, selon le principe de *limitation de la conservation* des données (cf fig.1, point 4), la durée de la conservation ne doit pas excéder ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Dans une démarche de conformité, il conviendra donc de déterminer une telle durée conforme aux finalités de la recherche et veiller à ce que celle-ci soit respectée : à l'expiration de cette durée, les données doivent être supprimées ou anonymisées. Pour répondre aux besoins particuliers de la recherche scientifique (notamment l'exigence de la reproductibilité de la recherche), le RGPD prévoit la possibilité de déroger à ce principe, sous condition de mettre en œuvre les mesures techniques et organisationnelles appropriées pour garantir la sécurité et la confidentialité des données. Toutefois, lorsque les mêmes finalités peuvent être atteintes par des données anonymisées, cette voie est à privilégier[4]. En parallèle (cf. fig1, cercle 2), il faudra toujours garantir la sécurité des données à caractère personnel à l'aide de mesures techniques ou organisationnelles appropriées (*intégrité et confidentialité*). À titre non limitatif, le RGPD (art. 32) cite plusieurs techniques permettant d'assurer la sécurité du traitement, notamment la pseudonymisation[5], fortement encouragée par le RGPD (y compris dans le régime des garanties et dérogations applicables aux traitements à des fins de recherche scientifique[6]). De plus, conformément au *principe de loyauté et de transparence*, le traitement doit être transparent à l'égard des personnes concernées (cf fig.1, cercle 3). Les articles 12 à 14 du RGPD imposent, respectivement, une certaine qualité de communication des informations, qui doit être *« concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples, en particulier pour toute information destinée spécifiquement à un enfant* », et une liste des informations qui doivent être communiquées.

En principe, depuis l'entrée en vigueur du RGPD, la mise en œuvre de la protection des données personnelles, telle que décrite ci-dessus, relève des responsables de traitement : le rôle de la CNIL consiste, le cas échéant, à contrôler cette conformité. Toutefois, certains traitements en matière de *Learning Analytics* peuvent présenter des risques élevés pour les droits et libertés des personnes physiques (surveillance systématique de leur comportement en ligne, données des mineurs...). La CNIL définit ainsi, sur la base de l'article 35 du RGPD, une liste des types d'opérations de traitement pour lesquelles il est nécessaire de réaliser une analyse d'impact relative à la protection des données (AIPD) [3], qui complète la liste de 9 critères issus des lignes directrices

[4] V. RGPD, art. 89.1

[5] Celle-ci est définie comme « le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable est d'ailleurs » (RGPD, art. 4.5)

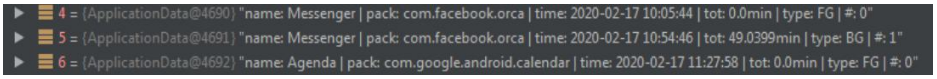
[6] RGPD, art. 89

du G29 [8]. Il convient alors, à l'aide d'une méthodologie adaptée[7] et en collaboration avec le DPO, d'apprécier la gravité et la vraisemblance des risques, ainsi que les mesures mises en place pour les prévenir. S'il apparait que le niveau de risque résiduel reste élevé malgré ces mesures, la CNIL doit être consultée avant la mise en œuvre du traitement.

3.2 Mise en œuvre dans un projet d'EMI en contexte scolaire

La mise en œuvre que nous décrivons est réalisée dans le cadre d'un projet de recherche qui s'inscrit dans le domaine de l'EMI (Éducation aux Médias et à l'Information) dans le contexte scolaire. La contribution par la recherche de ce projet porte sur l'éducation des adolescents à l'usage raisonné du smartphone pour réguler leur temps de sommeil et l'allocation de leur attention sur le temps scolaire. Le smartphone étant un équipement personnel, toujours disponible, nomade et connecté, il est propice aux usages personnels quotidiens, ce qui impacte le sommeil et l'attention des adolescents quand son utilisation n'est pas raisonnée. Le public ciblé est un public de lycéens entre 15 et 18 ans qui sont les plus exposés aux risques numériques.

Afin d'inciter les lycéens à un usage raisonné du smartphone, nous avons développé une application mobile dans le cadre de l'enseignement de l'EMI qui leur restitue leurs pratiques numériques [1] sous forme de tableaux de bord. La restitution est rendue possible grâce à la collecte et à l'analyse de leurs données d'interaction issues de l'utilisation du smartphone. La collecte de données est réalisée grâce à une application mobile qui « écoute » toutes les actions de l'utilisateur et transmet les données d'interaction sur un serveur sécurisé. Cette application est constituée de deux couches : une couche Application mobile à installer pour collecter les données d'interaction des adolescents, une couche Serveur caractérisée par une base de données qui stocke les données collectées et une API qui fait office de module de transmission des données entre l'application mobile et la base de données. Plusieurs données sont collectées, par exemple : les actions de marche/veille du smartphone, la consommation de données par application, les applications utilisées, les interactions avec les applications, les URLs visitées, la localisation, les notifications et réactions utilisateurs à ces notifications… À titre d'illustration, nous présentons dans la figure 2 un exemple de données brutes décrivant les interactions avec les applications ainsi que son format. Il s'agit de collecter des données décrivant les applications ouvertes par l'utilisateur et comment l'utilisateur interagit avec elles à travers l'attribut "action" : FG (foreground) quand une application passe en premier plan suite à l'interaction de l'utilisateur et BG (background) quand l'application passe en arrière-plan. Nous retrouvons également le nom de l'application, l'horodatage ainsi que la durée de l'interaction.



[7] La CNIL met à dispositions du public un logiciel open source (PIA : Privacy Impact Assessment), afin de faciliter la conduite et la formalisation d'AIPD.

Fig. 2. Exemples de données brutes décrivant les interactions avec les applications : Format données = {(application, package, horodatage, durée, action (FG | BG))

Dans cette section, nous présentons la mise en œuvre des orientations décrites dans la section 3.1 dans le projet d'EMI. Tous les principes évoqués et ces orientations ont été mis en œuvre mais nous ne pouvons pas tous les décrire. Nous nous focalisons sur certains critères ayant donné lieu à des fonctionnalités dans l'application d'EMI que nous pouvons illustrer.

3.2.1 Mise en œuvre du principe de transparence

Celle-ci prend la forme d'une **notice d'information** que nous avons rédigée et qui précise l'ensemble des éléments qui doivent être communiquées selon l'article 13 du RGPD (identification du responsable de traitement, les finalités du traitement, la base légale, les données traitées, les destinataires des données, les modalités d'exercice des droits des personnes, la durée de conservation,...). Conformément à l'article 12 du RGPD, cette notice a été intégrée dans l'application de collecte de telle manière qu'elle puisse être facilement accessible à l'utilisateur tout le temps. Elle se présente dans l'onglet « Données personnelles » de l'application (Figure 3).

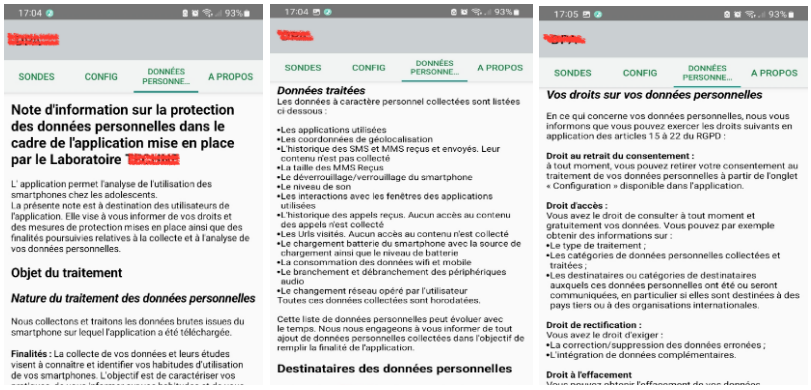


Fig. 3. Quelques extraits de la notice d'information intégrée dans l'application d'EMI

3.2.2 Mise en œuvre du principe de licéité : consentement

Le fondement de licéité retenu dans le cadre de cette recherche est le consentement. L'application de collecte a été façonnée de sorte à permettre l'expression d'une volonté libre, spécifique, *éclairée* et *univoque,* conformément à l'article 4.11 du RGPD [6]. Pour recueillir le consentement, nous avons mis en œuvre une interface listant les données collectées avec une case à cocher indiquant l'accord de l'utilisateur (cf. fig 4 à gauche) qui peut être décochée pour retirer le consentement à n'importe quel moment. Ce consentement est enregistré sous la forme d'un booléen dans une table de la base de données dédiée au consentement. Pour renforcer le caractère spécifique du consentement de l'utilisateur, nous lui avons également donné la possibilité de choisir les données qu'il autorise à collecter. Ces choix sont opérés par un système « d'interrupteurs » que l'utilisateur peut actionner ou non comme illustré dans la figure 4 à droite. Les

interrupteurs verts correspondent aux données pour lesquelles le consentement est accordé contrairement à celles en rouge.

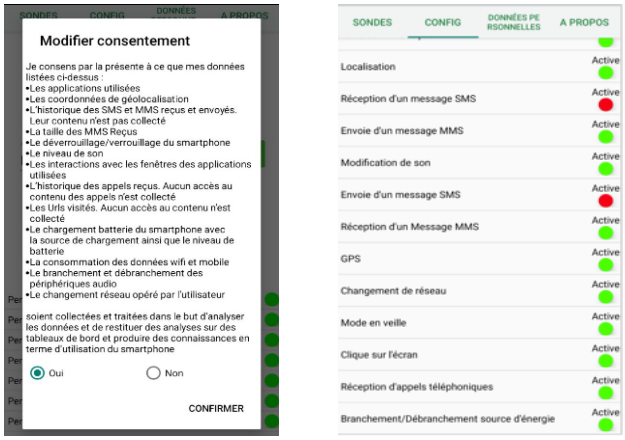


Fig. 4. Le consentement et son retrait (à gauche), sa personnalisation (à droite)

3.2.3 Mise en œuvre du principe de sécurité

Contrairement aux données anonymes, les données personnelles « pseudonymisée » continuent de caractériser une personne [7] en la désignant par un code plutôt que par un nom, comme vu précédemment. Dans notre contexte de suivi des pratiques numériques dans le temps, l'anonymisation n'est pas possible : nous optons donc pour des mesures techniques (script informatique) consistant à pseudonymiser les données de l'utilisateur et à lever cette pseudonymisation pour la restitution dans les tableaux de bord mais de manière totalement opaque et sans que l'on puisse identifier la personne dont les données sont collectées. Dans la figure 5, à droite est présenté un extrait les données de la table "usage des applications" qui décrit les interactions de l'utilisateur. Dans la structure de la table, la pseudonymisation est mise en œuvre à travers l'attribut (device_ref). Il s'agit de la donnée permettant de caractériser l'utilisateur par un code plutôt que par son nom ce qui ne permet pas de le rendre identifiable tout en préservant la possibilité de lui restituer l'analyse de ses données.

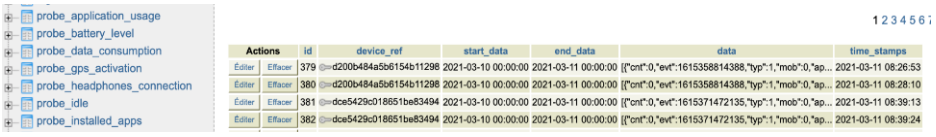


Fig. 5. Base de données et pseudonymisation

4 Conclusion

Dans cet article, nous avons présenté un guide pour la mise en conformité avec le RGPD des travaux de recherche portant sur la conception de dispositifs basés sur la collecte

de données à caractère personnel. Nous avons également donné quelques exemples de mise en œuvre dans le cadre de la conception d'un dispositif d'EMI en contexte scolaire. Cependant, cette mise en œuvre n'est pas présentée dans sa totalité conformément à ce que nous avons réalisé. En effet, il est difficile d'être exhaustif dans cet article car la prise en compte du RGPD dans notre recherche a été un processus long et très documenté. Pour de nombreuses recherches en EIAH, le RGPD apporte de nouvelles difficultés et de nouveaux défis : mise en conformité, acculturation à la donnée, lisibilité de certains métiers comme celui du DPO…Le RGPD et le droit du numérique deviennent des composantes incontournables dans la recherche en EIAH où des dispositifs sont conçus dans une logique de production et d'usages des données. Leur gestion précise, claire et sécurisée est une condition sine qua non pour instaurer un climat de confiance entre chercheurs et utilisateurs de dispositifs d'apprentissage. Les chercheurs en droit du numérique doivent ainsi avoir toute leur place en tant que partenaires des projets de recherche en EIAH contribuant ainsi à enrichir encore plus le caractère pluridisciplinaire du domaine.

Références

1. Aillerie, K.: Pratiques informationnelles informelles des adolescents (14-18 ans) sur le Web. Thèse en Sciences de l'information et de la communication. Université Paris 13 (2011).
2. CJUE, gr. ch., *OT c Vyriausioji tarnybinės etikos komisija*, C184/20 (source jurisprudentielle), 1 août 2022.
3. CNIL, https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-aipd [consulté en dernier le 27 décembre 2022].
4. CNIL, Délibération n° 2015-255, aff. JCDecaux. V. aussi, rejetant le recours, CE, 10e et 9e ch., 8 févr. 2017, n° 393714, JCDecaux c/ CNIL 20 (source jurisprudentielle),16 juillet 2015.
5. Dabbebi I, Iksal S., Gilliot JM, May M, Garlatti S: Towards Adaptive Dashboards for Learning Analytic: An Approach for Conceptual Design and implementation in (CSEDU 2017), , Porto, Portugal. pp.120-131, Apr 2017.
6. G29, Lignes directrices sur le consentement au sens du règlement 2016/679, 10 avril 2018.
7. G29, Avis n° 05/2014 sur les techniques d'anonymisation, WP 216, 10 avr. 2014.
8. G29, Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est « susceptible d'engendrer un risque élevé » aux fins du règlement (UE) 2016/679, 4 avr. 2017 (modifiées 4 oct. 2017), WP 248.
9. Les lignes directrices du Groupe de l'article 29 sur la transparence au sens du règlement (UE) 2016/679.
10. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi du 20 juin 2018 et par l'ordonnance du 12 décembre 2018 (source légale).
11. Pierrot L., Cerisier JF, El-Kechaï H., Ramírez S, Pottier L: Using a mixed analysis process to identify the students' digital practices. In EC-TEL 2017 Tallinn, Estonia (2017).
12. Règlement (UE) 2016/679 du Parlement Européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (source légale).