

12 janvier 2017

Synthèse de la journée

Julien Rossi (COSTECH – UTC)
Florent Pasquier (COSTECH – ESPE Paris)

Il y a presque un an, l'Espé de Paris (Université Paris-Sorbonne) organisait une journée d'études interrogeant les TICE (Technologies de l'information et de la communication pour l'enseignement) dans leur dimension « bienveillantes ». Mais est-ce encore le cas ? Quelles conséquences pourrait entraîner la généralisation du nouveau livret scolaire unique numérique (LSUN) en cours d'implantation dans le système scolaire du CP à la 3ème, puis plus tard probablement au delà jusqu'à l'université ? Dans la même veine, un article récent du *Monde* présentait un système de classe à distance d'une grande école madrilène dotée d'un dispositif de reconnaissance et de traitement en temps réel des émotions des étudiants, pour que l'enseignant ajuste son cours à leur degré d'attention, et sans que soit connu le devenir de ces données personnelles qui permettraient aussi à l'école d'extrapoler les capacités cognitives des utilisateurs par des traitements « big data » de leurs émotions (Miller, 2016).

La question de l'exploitation des données à caractère personnel à l'université est un sujet encore relativement peu abordé. Cela est particulièrement frappant en ce qui concerne la collecte et l'exploitation des données personnelles générés par les étudiants dans le cadre de relations d'enseignement utilisant de plus en plus des dispositifs numériques captant un nombre croissant de données étudiantes. La littérature scientifique, à ce jour, semble à peu près inexistante.

L'emploi de données personnelles pour des finalités de recherche scientifique a fait l'objet d'un peu plus d'attention. Des travaux ont notamment été réalisés sur la thématique de l'anonymat et des difficultés qui y sont liées (Ohm, 2010) (Sweeney, 2000) (Mascetti et al., 2013), en particulier en ce qui concerne les données de santé (Quinn, 2015). En plus de la formulation de divers projets scientifiques fondés notamment sur l'exploitation des traces du web (Boullier, 2015), il existe un effort critique de théorisation épistémologique et méthodologique important autour de la notion de trace numérique (Jeanneret, 2011) (Collomb, 2016).

Mais comme pour l'utilisation de données étudiantes dans la relation d'enseignement, les questions soulevées par l'exploitation de données personnelles à des fins de recherche sont loin d'avoir toutes reçu une réponse satisfaisante, et au-delà d'un cercle de spécialistes, elles ne font que rarement l'objet d'une réflexion institutionnelle au sein de la profession des enseignants-chercheurs.

En parallèle à ces travaux académiques, une communauté de professionnels de la protection des données personnelles, notamment constituée de correspondants informatique et libertés (CIL), a développé, en partenariat avec la Commission nationale de l'informatique et des libertés (CNIL), une grande expertise dans l'application au monde universitaire du droit des données personnelles. Ces professionnels se sont regroupés au sein de plusieurs réseaux, comme SupCIL, pour les CIL de l'enseignement supérieur et de la recherche, l'Association française des correspondants aux données personnelles (AFCDP) qui regroupe les CIL de France ou encore, au niveau international, l'International Association of Privacy Professionals (IAPP). Cette expertise nourrie de la pratique porte tant sur les données des personnels que sur les données étudiantes et les données traitées à des fins de recherche scientifique.

Le but de la journée d'étude du 12 janvier 2017 sur les données personnelles en milieu universitaire, organisée par le laboratoire COSTECH de l'Université de technologie de Compiègne à l'Institut des sciences de la communication (ISCC), était double :

- Sur le plan scientifique, il s'agissait de présenter l'état de la recherche sur le traitement de données personnelles à l'université ;
- Sur le plan institutionnel, il s'agissait de proposer un dialogue entre CIL et enseignants-chercheurs pour alimenter une réflexion en interne sur les questions méthodologiques, éthiques et juridiques liées à la protection des données personnelles en université.

Ces deux objectifs se déclinaient selon deux thématiques, la première abordée le matin et la seconde l'après-midi : les données personnelles dans la relation d'enseignement, et les données personnelles traitées à des finalités de recherche scientifique.

En raison d'un corpus de recherche restreint, en particulier sur le thème des données étudiantes, le fil conducteur de la journée était : « quelles questions se poser ? ». Cette question était à entendre tant sur le plan de la démarche éthique à adopter que sur le plan des sujets de recherche qu'il reste à aborder et que cette journée a aidé à déterminer.

Les notes prises au cours de la journée ont servi à proposer une première synthèse orale à la fin de la journée d'étude, laquelle a fait l'objet d'une discussion. En reprenant ces notes, nous avons compilé et structuré les principales questions qui, nous semble-t-il, ont émergé au cours de cette journée d'étude¹.

1) Une réflexion à construire sur les données personnelles dans la relation d'enseignement

A) Résumé des interventions

Le panel de la matinée était dédié à la collecte et à l'exploitation des données personnelles dans la relation pédagogique :

- 10h - 10h30 : Jérôme Valluy : Données personnelles et relation didactique... pour une critique de tri axiologique
- 10h30-11h : Célya Gruson-Daniel : Les coulisses d'une MOOC Factory : une plongée abrupte dans l'univers juridique et éthique
- 11h-11h30 : Rafia Bérenguier et Victor Larger : Mise en conformité des activités pédagogiques et administratives d'une université. Le cas de l'UTC

Jérôme Valluy a montré la difficulté qu'il y avait à réfléchir sur une telle absence de corpus théorique et scientifique sur le sujet. Ce vide dans la littérature est d'autant plus étonnant que se multiplient des projets pédagogiques numériques aboutissant à une captation grandissante de données personnelles étudiantes. En se basant sur son expérience d'enseignant-chercheur, et en analysant la structure de l'environnement numérique de travail de son université (ENT similaire à celui des autres universités) il montre que les outils mis à disposition aboutissent à la captation et à la centralisation d'un grand nombre de données étudiantes, peut-être au-delà de « l'intérêt légitime de l'enseignant » (au sens d'un intérêt d'efficacité didactique orientée par ce qui est utile et indispensable à l'apprentissage des étudiant-e-s). Cette centralisation, dans un contexte d'affaires médiatisées d'externalisations involontaires ou délibérées de données estudiantines, favorise un sentiment de méfiance de la part des usagers, et une sous-utilisation systémique des outils numériques, à laquelle il peut être remédié en partie par l'emploi d'outils externes (comme ceux de Framasoft) dispersés sur plusieurs serveurs éloignant les données étudiantes des divers contrôles et

¹ Ainsi, ce compte-rendu n'engage que ses auteurs et non l'ensemble des participants à la journée d'étude

réutilisations potentielles, ultérieures et extérieures, à des fins (par exemple commerciales ou politiques) autres que proprement universitaires.

Célya Gruson-Daniel, en se basant sur son expérience au sein d'une MOOC-Factory, a montré comment et pourquoi les concepteurs de cours en ligne utilisaient les données des apprenants, notamment à des finalités de recherche. Cela posait la question de la frontière floue, dans ces dispositifs, entre les données collectées à des finalités pédagogiques, et celles collectées à des finalités de recherche. Par ailleurs, Célya Gruson-Daniel a montré que l'intégration des contraintes juridiques était souvent perçue comme une intrusion, voire un fardeau, dans le travail et le métier des concepteurs de MOOC.

Enfin, Rafia Bérenguier et Victor Larger du réseau SupCIL ont présenté le récit de la mise en conformité d'une université, en se fondant sur l'expérience de l'Université de technologie de Compiègne et de l'Université Paris Descartes. Cela a permis de faire le lien avec les contraintes juridiques et de présenter le rôle des CIL, et de souligner certaines questions qui sont encore juridiquement en suspens. Par exemple, cela a été l'occasion de revenir sur la discussion autour de la qualification d'examens en ligne de « téléservices » au sens de l'article 27-II sous 4) de la loi Informatique et Libertés et de l'ordonnance « téléservices » de 2005, qui aboutirait à saisir systématiquement la CNIL pour avis avant la mise en place de tels services.

B) Questions dégagées

- Les enseignants sont en situation de domination objective et de pouvoir vis-à-vis des étudiants. La licéité d'un traitement de données personnelles que l'enseignant demande aux étudiants de fournir ne saurait donc être fondé sur le seul consentement de ces derniers. Le traitement ne saurait donc être fondé que sur l'intérêt légitime de l'enseignant ou, dans les établissements publics, sur le fait qu'il s'agit de l'exercice d'une mission de service public (art.7 de la loi Informatique et Libertés). Alors, quelles sont les limites de l'intérêt légitime de l'enseignant à collecter des données sur ses étudiants ?
- Le paradoxe de la vie privée est désormais bien documenté. Bien des gens affirment être soucieux de leur vie privée sans pour autant montrer des comportements en cohérence avec cet objectif (Acquisti et Gross, 2006). Ainsi, la question a été posée pendant les discussions de savoir si les étudiants avaient bel et bien un intérêt pour leurs données personnelles, dès lors qu'ils utilisaient massivement Facebook et y laissait un grand nombre de données personnelles, voire de données sensibles. Une piste de réponse apparue au cours des discussions pourrait être la notion d'éloignement. En effet, si les étudiants sont peu enclins à laisser leurs données à l'université ou au professeur, c'est parce que l'autorité collectant ces données est proche, et peut prendre des décisions affectant directement les étudiants sur la base de ces données. A l'inverse, Facebook est perçu comme un responsable de traitement lointain, désincarné, et ayant finalement peu d'influence sur la vie des étudiants qui l'utilisent. Cette notion d'éloignement de la personne concernée par rapport à l'autorité collectant les données peut être une piste de recherche intéressante à creuser pour résoudre ce fameux paradoxe de la vie privée.
- La protection des données personnelles, au-delà du fait qu'il s'agit d'un droit fondamental, est-il un élément nécessaire pour garantir la confiance dans une relation pédagogique numériquement médiée ?
- Comment concilier technologies de l'information et de la communication pour l'enseignement (TICE) et éthique ? La dispersion des données pour éviter leur centralisation

et éviter les recoupements comme l'exploitation commerciale, comme proposée par Jérôme Valluy, est-elle une piste de solution ?

- Comment, d'ailleurs, limiter le risque de diffusion hors-contexte des données étudiantes ? Ainsi, une note intermédiaire n'a pas de sens si elle n'est pas lue en combinaison ses résultats dans le temps et dans d'autres matières, mais peut, prise hors de contexte, être utilisée pour lui nuire.
- D'ailleurs, lesquelles de ces données doivent être exclues des principes de l'*open data* ? Parmi les données étudiantes, lesquelles ont vocation à être protégées, et lesquelles relèvent de travaux scientifiques (thèses, mémoire) pouvant voire devant être publiés ?
- Dans le cadre des MOOC, quelle est la limite entre les finalités pédagogiques et les finalités de recherche pour lesquelles les données sont collectées ? Célya Gruson-Daniel a montré la porosité entre ces deux notions. Dès lors, quel régime juridique s'applique à ces données, puisque les données exploitées à des fins de recherche bénéficient d'un cadre juridique spécifique ?

2) Les données personnelles traitées pour des finalités de recherche scientifique

A) Résumé des interventions

Le panel de l'après-midi était dédié à la question de l'exploitation de données personnelles à des finalités de recherche scientifique :

- 13h10 - 14h : Les humanités numériques face aux données personnelles (Panel du laboratoire COSTECH avec Jean-Edouard Bigot, Cléo Collomb, Julien Rossi et Eglantine Schmitt)
- 14h20 - 15h30 : L'étude du politique et l'enjeu de la protection des données (discutante : Francesca Musiani)
 - 14h20-14h45 : Julien Boyadjian : La science politique face aux enjeux du « big data » et de la protection des données personnelles sur Internet
 - 14h45-15h10 : Kaja Skowronska : Entre responsabilité éthique et contrainte formelle – le traitement des données confiées aux chercheurs lors des enquêtes qualitatives
- 15h45 - 16h30 : La tension entre progrès de la recherche et vie privée peut-elle être dépassée ? Le cas des données de santé. Panel du laboratoire LSTS de Bruxelles , avec Paul Quinn, Pedro Cristobal Bocos et Eugenio Mantovani
- 16h45 - 17h30 : Intervention du Correspondant informatique et libertés du CNRS (présentation faite par Emilie Masson)

Jean-Edouard Bigot, Cléo Collomb, Julien Rossi et Eglantine Schmitt ont montré, après les avoir définies, comment les traces numériques étaient théorisées et exploitées en sciences humaines et sociales (SHS). Ce panel a également été l'occasion de débattre des discussions autour de la définition des notions de « trace » et de « donnée personnelle ». Si ce n'est pas toujours la personne derrière la trace qui intéresse le chercheur en SHS, ils ont montré que les traces numériques tombaient – sauf rares exceptions – dans la catégorie juridique des données à caractère personnel, malgré des discours qui tendent parfois à le faire oublier. Cette qualification juridique impose au chercheur un cadre juridique qui peut être contraignant pour sa recherche, mais qui lui-même soulève un grand nombre de questions.

Julien Boyadjian a cherché à montrer l'intérêt scientifique de travailler sur des données personnelles disponibles sur le web, tout en soulevant les nombreuses difficultés méthodologiques, éthiques et finalement juridiques qu'un tel travail implique. Kaja Skowronska a illustré quant à elle la façon dont le droit des données personnelles proposait aux chercheurs des outils qui leur permette de protéger la confidentialité des informations parfois sensibles que les enquêtés leur transmettent. Cet emploi du droit de la protection des données peut venir modifier le rapport de force tripartite entre le chercheur, l'enquêté et le cadre institutionnel dans lequel évolue le chercheur.

Paul Quinn, Pedro Cristobal Bocos et Eugenio Mantovani ont montré quant à eux les limites de l'anonymisation comme solution au dilemme entre protection des données et de la vie privée d'une part, et l'intérêt public représenté par l'avancée de la recherche scientifique d'autre part. Ils rappellent l'existence, notamment dans le nouveau Règlement général de protection des données (RGPD) (2016/679/UE) d'un cadre juridique permettant aux traitements à des fins de recherche scientifique d'échapper à certaines contraintes, comme l'information des personnes concernées ou la limitation des finalités. Ce cadre juridique est soumis à des conditions pouvant être précisées par le droit national malgré l'adoption du règlement censé proposer une harmonisation plus poussée que l'actuelle directive 95/46/CE, mais les contours de ces conditions nationales sont aujourd'hui peu connues et suscitent de fortes interrogations.

Enfin, Emilie Masson a présenté pour le CIL du CNRS le détail du cadre juridique auquel les chercheurs doivent se conformer. Elle a notamment étudié les dispositions du RGPD, en délimitant tout comme le panel précédant un certain nombre d'incertitudes.

B) Questions soulevées

Contrairement à la question des données personnelles dans la relation d'enseignement, la question des données de recherche a déjà fait l'objet d'un certain nombre de travaux. Nous nous centrerons ici sur les questions dont les discussions ont permis mettre en lumière un manque dans la littérature académique ou institutionnelle.

- Francesca Musiani, Julien Boyadjian et Kaja Skowronska, ont montré dans les discussions la relation étroite entre le droit de la protection des données et des réflexions plus anciennes sur la relation entre le chercheur et ses enquêtés. Cette relation pose des questions d'ordre éthique : pour qui fait-on de la recherche ? Elle soulève également des problèmes méthodologiques renforcés par les principes juridiques de la protection des données. Ceci est en particulier le cas du consentement, qui introduit un biais de sélection, ou de l'information des personnes concernées, qui biaise les comportements et donc les résultats des enquêtes. Mais ce droit fournit aussi des outils à la mise en pratique de réflexions éthiques antérieures dont Kaja Skowronska a rappelé la place dans les SHS.
- Le RGPD prévoit, notamment à son article 89, un statut dérogatoire pour les traitements ayant pour finalité la recherche scientifique. Combiné à d'autres articles du même règlement, les chercheurs bénéficient potentiellement de larges dérogations aux principes de la protection des données. Sera-t-il cependant possible d'en bénéficier, et sous quelles conditions ? Selon ce qui est ressorti des présentations et des discussions, les autorités de protection des données semblent peu enclines à permettre aux chercheurs de bénéficier facilement de ces dérogations, et continuent à favoriser soit l'anonymisation des données dès la collecte, soit l'application du régime général de protection des données à la recherche.
- Face à cela, convient-il de défendre ce statut dérogatoire ? Ce statut est légitimé en large partie par les attentes fondées dans ce que le *big data* et l'exploitation massive de traces

numériques promettent d'apporter à la recherche. Or, ces attentes sont en large partie construites sur des mythes peu scientifiques, voire idéologiques. Dès lors, comment fonder scientifiquement et rigoureusement la justification de ce statut dérogatoire ? La présentation que Julien Boyadjian a faite lors de la journée d'étude nous semble fournir quelques pistes intéressantes à explorer pour entamer la réflexion à ce sujet.

- Enfin, les contours mêmes du cadre juridique pour les traitements à des fins de recherche, prévu par le RGPD, sont encore à préciser, par la jurisprudence et la doctrine des autorités de protection des données comme la CNIL, mais aussi et surtout par le droit national. Dès lors, des différences de régime juridique entre Etats membres de l'Union sont à prévoir en la matière. Ces différences s'ajoutent au fait que ce statut dérogatoire ne prévoit pas d'exception au principe selon lequel les données personnelles ne peuvent être transmises, sauf exceptions limitativement énumérées, vers des pays en dehors de l'Union européenne. Comment s'y retrouver dans le cadre de projet européen, ou internationaux ? Dans sa discussion du panel politiste, Francesca Musiani a insisté sur cet aspect en tirant des exemples de son expérience au sein de projets de recherche européens. Sur ces sujets, un avis du Groupe de travail de l'Article 29, sur le modèle du *Guide informatique et libertés dans l'enseignement supérieur* que la CNIL a publié en 2011, serait le bienvenu.

3) Questions communes aux deux thématiques

Bien que l'université regroupe à la fois des traitements de données étudiantes et des traitements de données personnelles exploitées pour des finalités de recherche scientifique, le lien entre les deux thématiques peut paraître artificiel au-delà du fait qu'il s'agit des deux volets que les enseignants-chercheurs ont à traiter dans le cadre de leur activité professionnelle. La journée d'étude du 12 janvier a cependant permis de relever un certain nombre de problématiques communes, que nous avons choisi de lister à part.

A) Protection des données et enjeux de pouvoir

- Beaucoup de discussions ont tourné autour de questions d'enjeux de pouvoir. Le droit de la protection des données crée-t-il un nouveau lieu de pouvoir au sein des universités, à la façon des *Institutional Review Boards* anglo-saxons ? Existe-t-il un risque que les présidents d'universités, ou les directeurs d'unités, sous couvert de protection des données, s'immisce dans l'autonomie des chercheurs ? La protection des données impactera-t-elle de façon égale tous les projets de recherche, indépendamment de leur caractère orthodoxe ou hétérodoxe au sein des disciplines académiques ? Ces questions semblent créer une forme de méfiance qui peut compliquer le dialogue entre CIL et enseignants-chercheurs.
- En particulier, une question soulevée régulièrement est : qui jugera de la nécessité scientifique d'utiliser telle ou telle donnée ? Ou de sa nécessité pédagogique ? Comment éviter l'instrumentalisation de la protection des données dans des enjeux de pouvoir, et protéger ainsi notamment l'indépendance du CIL face à ces enjeux de pouvoir ?
- La présentation de Kaja Skowronska a montré par ailleurs que les outils contenus dans le droit de la protection des données pouvait avoir un effet de modification de relations de pouvoir dans le champ universitaire. Cet effet n'a que peu ou pas été étudié en recherche mais semble pourtant sous-tendre, comme nous l'avons souligné dans les points précédents, un grand nombre de discussions.

B) Les questions liées à la sécurité des systèmes d'information

- La sécurité des données personnelles est un des principes de la protection des données, qui s'applique à la fois aux données étudiantes et aux données traitées à des fins de recherche. Dans les deux cas, les enseignants-chercheurs peuvent être amenés à protéger des données particulièrement sensibles. C'est le cas pour les données étudiantes lorsqu'ils enseignent dans des pays dictatoriaux, ou encore lorsqu'ils traitent des données de santé sur leurs enquêtés.
- Une absence de formation et parfois d'équipement sécurisé caractérise cependant l'enseignement supérieur et la recherche. Comment, dans un contexte de restrictions budgétaires, assurer la formation des personnels et la mise à disposition d'un équipement sécurisé ? Quelles bonnes pratiques diffuser ?
- Ces questions ont été très peu abordées, par manque de temps et pour recentrer les enjeux de la journée d'étude sur les enjeux spécifiques de la protection des données. Ils ont néanmoins fait surface régulièrement au cours des débats, tant dans la matinée qu'au cours de l'après-midi.

Conclusion

Le panorama ci-dessus de questions soulevées pendant la journée d'étude du 12 janvier n'est pas exhaustif. Il exclut notamment un grand nombre de questions ayant déjà fait l'objet de travaux académiques, en particulier pour ce qui concerne les recherches portant sur les techniques d'anonymisation. Ce parti pris a permis d'écourter un peu cette liste déjà longue pour mettre en lumière les nombreuses zones d'ombre tant de la recherche que, parfois, des discussions institutionnelles parmi les enseignants-chercheurs et entre ces derniers et la communauté des professionnels de la protection des données.

Les autres documents mis en ligne sur la page web de la conférence permettront à chacun de creuser les différentes questions soulevées par les multiples intervenants, que nous profitons de ces quelques lignes pour remercier encore de leur participation.

Références bibliographiques

ACQUISTI A., GROSS R., 2006, « Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook », *Proceedings of the 6th International Conference on Privacy Enhancing Technologies*, p. 36-58.

BOULLIER D., 2015, « Les sciences sociales face aux traces du Big Data », *Revue française de science politique*, 5-6, p. 805-828.

COLLOMB C., 2016, « Pour un concept technologique de trace numérique », *Azimuth. Philosophical Coordinates in Modern and Contemporary Age*. IV, 7, p. 37-60

JEANNERET Y., 2011, « Complexité de la notion de trace », dans GALINON-MÉLÉNEC B. (dir.), *L'homme-trace : perspectives anthropologiques des traces contemporaines*, Paris, CNRS-éditions, p. 59-86.

MASCETTI S., MONREALE A., RICCI A., GERINO A., 2013, « Anonymity : A Comparison Between the Legal and Computer Science Perspectives », dans *European Data Protection: Coming of Age*, Dordrecht, Springer, p. 85-115.

MILLER M., 2016, « A Madrid, des étudiants sous l'œil du big data », *Le Monde.fr*, 1 novembre 2016, http://www.lemonde.fr/campus/article/2016/11/01/a-madrid-des-etudiants-sous-l-il-de-big-data_5023668_4401467.html

OHM P., 2010, « Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization », *UCLA Law Review*, 57, p. 1701-1777.

PASQUIER F. 2017, « Journée d'études les tice bienveillantes », dans le programme hébergé "Humanisme, éducation, technologie et faits sociaux", <http://www.reseau-terra.eu/rubrique309.html>

QUINN P., 2015, « The Use of Anonymisation Techniques to Allow Patient Data to be Used for Research Processes – A Reflection on the Recent Article 29 Working Party Opinion », dans *Fifth European Conference on Health Law, Book of Abstracts*, p. 23.

SWEENEY L., 2000, « Uniquement of Simple Demographics in the U.S. Population », *Laboratory for International Data Privacy, Working Paper LIDAP-WP4*.