



Enjeux, usages, éthiques et droit du Big Data

Vendredi 25 mars 2016

à 9h30

Ecole de Journalisme et de Communication Amphi A

O u v e r t u r e		
9 :30	Alexandre Joux , Directeur de l'EJCAM Françoise Bernard , Directrice de l'IRSIC Cristinel Diaconu , CPPM, Animateur du groupe d'animation PR2I Big Data	
9h40	Fidelia Ibekwe-SanJuan et Françoise Bernard — IRSIC <i>Enjeux du big Data pour les SHS et les SIC</i>	F. Ibekwe-SanJuan
10h10	Cédric Motte , Journaliste <i>La difficile relation entre les journalistes et les données</i>	
11h	P A U S E	
11h10	Alexandre Joux et Marc Bassoni — IRSIC <i>Le journalisme à l'heure des « big data ». Questions et perspectives.</i>	F. Bernard
11h40	Florence Andreacola — Centre Norbert Elias (EHESS), Univ. Avignon. <i>Expérience de visite et données en ligne : accès, formats et finalités pour l'institution muséale</i>	
12h10	Florian Masse, Sébastien Oliveau, Samuel Carpentier-Postel, Frédéric Audard, Lionel Kieffer — ESPACE UMR 7300 <i>Des Big Data à la place des enquêtes ? Enjeux scientifiques et éthiques pour l'étude des mobilités quotidiennes.</i>	
12h40	B U F F E T S U R P L A C E	
14h	Nahela EL BIAD — CERIC <i>Le droit, le Big data et le traitement des données de santé.</i>	J-C Dufour
14h30	Marc Tanti — INSERM-IRD-UMR912-SESSTIM <i>L'exploitation des données du web social : une nouvelle approche méthodologique et de nouveaux enjeux en santé publique</i>	
15h00	Audrey Bonjour — IRSIC Vincent Meyer et Elise Daragon - I3M, Univ Sophia-Antipolis <i>Les technologies numériques dans les établissements et services sociaux et médico-sociaux : Enjeux, Usages, Ethiques et Droit du Big et Small Data</i>	
15h30	Thierry Gobert — Univ. Perpignan <i>Le « micro big data » : projections et ruptures en C2i</i>	
16h	D E B A T S & C L O T U R E Françoise Bernard (IRSIC), Cristinel Diaconu (CPPM), Jean-Charles Dufour (SESSTIM), Fidelia Ibekwe-SanJuan (IRSIC)	

Résumé des présentations

1. Fidelia Ibekwe-SanJuan et Françoise Bernard — IRSIC

Enjeux du big Data pour les SHS et les SIC

Résumé : Un bref état des lieux des questions et enjeux suscités par le phénomène de Big Data pour les SHS et pour les SIC en particulier.

2. Cédric Motte, Journaliste

La difficile relation entre les journalistes et les données

Résumé : En France les médias sont encore timides dans l'utilisation de jeux de données. Si certains jeux sont utilisables relativement facilement, tels ceux estampillés open data, se pose toujours la question « Comment sont fabriquées ces données ? »

Quelle confiance un journaliste peut-il accorder à un fichier excel ou à une table de données ? Quels sont les moyens mis à sa disposition pour les vérifier ? Et comment les utiliser ensuite ? Nous verrons une suite d'exemples qui illustrent et tentent de répondre à ces questions.

3. Alexandre Joux et Marc Bassoni — IRSIC

Le journalisme à l'heure des « big data ». Questions et perspectives.

Résumé : Confronté aux « big data », le journalisme doit-il simplement s'adapter à un nouveau dispositif technique ? Si oui, comment et en quels termes ? *A contrario*, n'est-il pas plutôt à l'orée d'une mutation en profondeur ? Si tel est bien le cas, quelles sont les perspectives et quelles sont les questions-clés associées à un tel basculement paradigmatique ? En croisant la réflexion sur les Big data avec la recherche sur le journalisme de données, quatre questions-clés sont abordées. Ces questions sont les suivantes :

- Un nouveau rapport au « terrain » et aux sources d'information ?
- L'assomption d'un journalisme « tiré par la demande » ?
- Une personnalisation accrue des contenus proposés ?
- Une gestion renouvelée des emplois et des qualifications au sein des rédactions

4. Florence Andreacola — Centre Norbert Elias (EHESS), Univ. Avignon.

Expérience de visite et données en ligne : accès, formats et finalités pour l'institution muséale.

Résumé : À partir d'une recherche de terrain menée au sein d'un musée de société français sur les pratiques numérique des ses visiteurs, nous avons traité, de façon croisée, des données issues du web et de données issues de protocoles classiques en sciences humaines e sociales comme l'entretien ou le questionnaire. Cette recherche interdisciplinaire nous a permis de mener une réflexion sur les caractères « explicites » et « implicites » des traces produites par l'activité du visiteur. Elle nous donne aussi la possibilité d'explorer les nouvelles fonctions de fournisseur de données que peuvent revêtir les institutions muséales. Cette intervention a pour objectif de présenter de façon analytique ce qui peut conditionner la production de données relatives aux usages des outils numériques et d'Internet dans le cadre d'une pratique de visite d'un musée. Ce travail met en évidence des logiques d'accès aux données et des logiques de formatage des informations collectées. Il nous permet, enfin, de faire émerger les différentes finalités assignées à ces données et qui coexistent par leur existence et leur traitement.

5. Florian Masse Sébastien Oliveau, Samuel Carpentier-Postel, Frédéric Audard, Lionel Kieffer —
ESPACE UMR 7300

Des Big Data à la place des enquêtes ? Enjeux scientifiques et éthiques pour l'étude des mobilités quotidiennes.

Résumé : Le développement récent de nombreuses technologies mobilisant la géolocalisation (par GPS, GSM, WiFi ou encore adresse IP) apparaît aujourd'hui comme une opportunité nouvelle pour documenter la question des déplacements des individus en ouvrant des possibilités que les enquêtes classiques ne peuvent offrir, notamment l'exhaustivité et le temps réel (Batty, 2012, Audard *et al.* 2014). La disponibilité de ces données constitue-t-elle pour autant une panacée ? Bien évident, non. Si les outils technologiques pourraient *a priori* fournir une large part des informations disponibles dans les enquêtes - et pour un certain nombre d'entre elles de manière bien plus fine (itinéraires notamment) - leur exploitation effective n'est pas compatible avec une notion essentielle autant dans le champ des enquêtes que dans celui des *Big Data* : le respect de l'anonymat. La conséquence concrète de cette nécessaire protection des données individuelles est, par exemple, de ne pas rattacher les caractéristiques sociodémographiques (âge, statut professionnel, etc.) aux données de déplacement à l'échelle individuelle (Ollion Boelaert, 2015). Or, en termes d'analyse ou d'aménagement du territoire, les enjeux reposent justement sur la connaissance des déterminants individuels des déplacements. L'intérêt de ces données est de ce point de vue fortement diminué. Entre questions méthodologiques (biais de sélection, incomplétude, etc.) et éthiques (anonymat, vie privée) les *Big Data* suscitent aujourd'hui autant d'espoirs que de questions. Nous nous proposons dans cette communication de revenir sur ces questions en les illustrant par une comparaison entre Enquête Ménages-Déplacements (EMD) et « traces » fournies par Orange.

6. Nahela EL BIAD — CERIC

Le droit, le Big data et le traitement des données de santé.

Résumé : Le développement des nouvelles technologies de l'information et de la communication a généré et générera un accroissement tout à fait exponentiel de ces données de santé. Par le biais des dossiers médicaux informatisés, le *quantified-self* ou encore l'implantation de dispositifs médicaux destinés à mesurer l'état de santé de patients, un nombre important de données est en effet émis et collecté puis stocké durant tout le parcours de soin d'un individu. Ces quantités massives de données numérisées, stockées et archivées auprès d'hébergeurs ont vocation, non seulement à réduire les coûts de santé, mais également à permettre un meilleur suivi de santé du patient avec l'ouverture d'une voie vers une médecine personnalisée et plus adaptée aux besoins de chacun. La gestion de ces données massives est également un important levier pour une meilleure appréhension des maladies, du commerce médicamenteux et du traitement des patients.

Le résultat est un renouvellement en marche de la relation du patient à ses données personnelles. En effet, dans les systèmes évoqués, les données sensibles que représentent les données de santé ont vocation à être partagées et ce, même si elles relèvent exclusivement de la vie privée de la personne. L'Open Data en santé va encore faire évoluer la relation au nom des possibilités d'amélioration du service rendu au patient.

Pour encadrer ce transfert de données de santé, quel est le cadre juridique actuel ? Y'a-t-il des évolutions envisagées ? En l'état actuel, l'échange et le partage des données de santé est fondé sur le respect du secret professionnel tel que défini au sein du Code pénal. Mais ce dogme du secret professionnel suffit-il ? Quelles sont les nouveautés prévues notamment dans le cadre de la loi Santé et du projet de Règlement ? Quels sont les risques du Big data sur la protection des données personnelles du patient ? Comment déterminer qui peut avoir accès à ces données et dans quelles conditions ? Quelle est la place du droit pour conjuguer l'intérêt collectif de diffusion des données de santé et la protection des personnes ? Comment se définirait la responsabilité si les données d'un patient étaient rendues publiques sans son accord ? Quelle est la place du consentement dans la mise en place du Big Data ? Enfin, quelles sont les évolutions possibles du droit pour parvenir à un niveau satisfaisant de sécurité des données de santé tout en promouvant ce nouveau mode de gestion des données de santé ?

7. Marc Tanti — INSERM-IRD-UMR912-SESSTIM

L'exploitation des données du web social : une nouvelle approche méthodologique et de nouveaux enjeux en santé publique.

Résumé : Les dernières avancées des Technologies de l'Information et de la Communication ont modifié en profondeur de nombreuses activités dans la sphère publique et dans la sphère privée entraînant des bouleversements de société. Notamment, avec l'émergence du Web 2.0, les modes de communications entre individus s'en sont retrouvés totalement révolutionnés. Ainsi, le nombre de données échangées sur les réseaux sociaux par les internautes s'en sont vu augmenter de manière exponentielle entraînant une pléthore d'informations dont certaines à très forte valeur ajoutée, mais aujourd'hui quasiment inexploitées en terme stratégique par les décideurs. Par exemple, sur les forums de discussion, les internautes s'échangent en temps quasi réel des informations sur leur santé ou sur leur préoccupations médicales. Ces données stratégiques constituent aujourd'hui pour la santé publique des « pépites » d'informations, quasiment inexploitées par les décideurs publics pour résoudre ou comprendre des problématiques de santé nationales. L'exploitation de ces mégadonnées permettrait ainsi de détecter l'émergence d'événements sanitaires inhabituels ou des tendances épidémiques. Elle permettrait à la société et ses décideurs d'anticiper des dynamiques de changements et certaines crises sanitaires. Elle permettrait de comprendre l'émergence de certaines préoccupations de santé pour pouvoir réagir et proposer des innovations.

Dans cet article, nous présentons trois projets innovants développés au sein de l'UMR 912 SESSTIM qui ont pour objectif d'exploiter les conversations du grand public sur les réseaux sociaux afin de mieux comprendre certaines problématiques de santé publique ou certaines controverses.

8. Audrey Bonjour — IRSIC

Les technologies numériques dans les établissements et services sociaux et médico-sociaux : Enjeux, Usages, Ethiques et Droit du Big et Small Data.

Résumé : Dans cette communication, nous présenterons un projet de recherche en cours « Les technologies NUMériques au SERVICE des USagers dans les établissements sociaux et médico-sociaux » (NUSERUS, 2015-2018) qui ouvrent bien des questions en termes d'enjeux, d'usages, d'éthiques et de droit du big et du small data. La problématique des données, de la collecte des données et de leur mode de traitement impacte aujourd'hui de manière croissante (et pressante) ce type d'organisation particulier que sont les établissements et services sociaux et médico-sociaux (ESSMS) qui possèdent une fonction de lien social forte. Par ailleurs, on sait que les technologies numériques en travail social permettent ou permettront dans un proche avenir, entre autres, de concevoir et de réaliser de nombreuses activités jusqu'alors accomplies dans des échanges interpersonnels qui sont le fondement de toute action dans ce champ ; ceci dans tous les secteurs des prises en charge et pour tous les publics : fragilisés, en situation de handicap ou marginalisés de la petite enfance au grand âge. Volontairement ou non, les pratiques professionnelles en travail social sont impactées par ces technologies.

Pourtant présentés il y a encore peu comme des milieux clos/cloisonnés voire des institutions totales au sens d'Erving Goffman, ces établissements font face à une demande sociale des usagers, celle d'accéder à l'information et par là même aux technologies et services numériques afférents. Cet enjeu d'un service public de l'information s'inscrit dans le contexte de l'évaluation des pratiques professionnelles (interne puis externe) depuis la loi 2002-2 rénovant l'action sociale et médico-sociale avec pour corollaire ce que d'aucuns nomment l'informatisation du travail social. Suivant les orientations et stratégies d'établissements, ces derniers se sont emparés différemment de la question des technologies, certains n'ayant pas ouvert la boîte de pandore, préférant résister à toute forme de techno-pratiques professionnelles.

Pour sérier et qualifier ces usages et les pistes technologiques d'avenir à développer dans les établissements et services sociaux et médico-sociaux (ESSMS) comme les coopérations souhaitables entre les professionnels du social, de la santé, les chercheurs et les développeurs des technologies numériques, nous menons sur trois années (2015-2018) une RA dans huit établissements. L'enjeu et l'originalité du projet est d'analyser les dispositifs sociotechniques

actuels et en devenir au service de l'utilisateur, ceci dans une perspective de formation *intra muros* des professionnels.

Nous présenterons ici la mise en place de cette recherche-action formation et les premiers résultats de nos observations, notamment en lien avec la problématique des données et de l'accès à l'information.

9. Thierry Gobert — Univ. Perpignan

Le « micro big data » : projections et ruptures en C2i

Résumé : Les Français sont méfiants vis-à-vis du Big Data (Levy, Gérard, Bartoli, 2016). Cette défiance serait notamment alimentée par la sphère politique et une méconnaissance du phénomène (*ibid.*). Le président du Conseil National du Numérique met ainsi en garde « contre l'automatisation à outrance » tout en y voyant « un potentiel infini » (Mahjoubi, 2016). L'un des chantiers à venir concernerait « les données fermées et peu utilisées que détient l'enseignement supérieur » (Ecoter, 2016), une institution qui formerait depuis vingt ans à l'analyse des données (Autemayou, 2014) sans les exploiter bien qu'elle pourrait « industrialiser la personnalisation de l'éducation » (Amiel, 2013) et accéder ainsi à de revenus innovants.

Pour autant, l'université négocie déjà un virage sans précédent dans son histoire. Elle s'adapte à l'avènement des sites sociaux, aux ressources pédagogiques concurrentes et aux outils « non payants » qui fonctionnent sur la base d'accès répétés à des données personnelles contre l'utilisation de fonctionnalités logicielles et la consultation de contenus. Les habitudes de consommation, la maîtrise d'une interface, la pression sociale et la responsabilité concourent à une fidélité aux usages (Gobert, 2016). La perspective d'un « fichage » invisible et la référence à Big Brother ne semblent pas inciter les apprenants à changer. Par exemple, ils emploient toujours aussi peu le courriel des Espaces Numériques de Travail (ENT) réputé exempts d'exploitation des informations privées.

Dans le contexte de défiance vis-à-vis de l'exploitation des données massives, nous avons souhaité analyser le différentiel observé entre l'utilisation de services GAFa connus pour leurs pratiques de Big Data et le sous-emploi d'outils relativement semblables implémentés dans les ENT sécurisés. Cette réflexion fait suite à un travail réalisé à propos d'un différentiel de même nature concernant l'utilisation d'outils socio-numériques privés pour gérer un diplôme plutôt que la plateforme pédagogique institutionnelle (Gobert, *ibid.*). Cette fois, la population cible est constituée de deux groupes de C2i niveau 2 (master). Les 54 apprenants sont invités à participer, pendant 7 séances, à une sensibilisation sur la « responsabilité et le management » liés au numérique avant de réaliser un dispositif en ligne évoquant les thématiques abordées.

Les premiers résultats font état d'une méconnaissance identique à l'analyse de l'étude Harris mais les effets de la recherche-action, « destinée à faire évoluer les usages par l'apport d'informations en prise directe sur les pratiques des acteurs » (Hess, 1983, p. 9) ont eu partiellement raison de cette méconnaissance. Cela dit, la confiance dans le big data n'a pas augmenté et les pratiques ont peu évolué. Le travail souligne l'influence de la projection faite sur les organisations d'une activité personnelle de « micro big data » (sans algorithmes) et ses impacts sur les illusions de contrôle (Gobert, 2015) et de compétences (Moser, 1992), fortement convoqués dans les pratiques et usages en lien avec le numérique.