

Appel à projets IDEX Université Grenoble Alpes **Initiatives de Recherche Stratégiques (IRS) – Année 2017/2018**

Acronyme du projet	DUTY
Titre du projet	Transparence des Données Anonymes ou Anonymisées
Porteur du projet	Emilie HOAREAU

Description du projet

(5 pages)

Description du projet et des résultats attendus (5 pages maximum au total)

- Contexte scientifique et/ou technologique, objectifs du projet et positionnement sur les scènes locale, nationale et internationale (1 page)
- Programme scientifique / Méthodologie / Résultats escomptés (3 pages)
- Organisation du projet : échéancier, personnel impliqué et pourcentage du temps prévu, partenariats, utilisation des crédits (1 page)

Contexte scientifique et/ou technologique

Le 25 mai 2018 entrera en vigueur le Règlement Général Européen sur la Protection des Données (RGPD) qui vient renforcer les exigences concernant la gestion des données à caractère personnel. Si les données associées à une personne identifiée ou identifiable sont désormais soumises à un cadre juridique, tel n'est pas encore le cas des données anonymes ou anonymisées (DAA). Le volume des DAA, déjà très important, est amené à connaître une croissance exponentielle avec l'entrée en vigueur du RGPD. Or, les entreprises disposent actuellement de peu de points de repères pour réguler la façon dont elles mobilisent les DAA. Les cadres normatifs utilisés jusqu'à présent se heurtent à la digitalisation rapide des activités. Malgré les efforts déployés sur la sécurisation des données, l'anonymat ne garantit pas une protection absolue des personnes. En effet, les DAA peuvent être attentatoires si elles sont mobilisées à des fins déloyales ou discriminatoires. Pour cette raison, le besoin d'une réflexion éthique s'impose afin d'accompagner les actions à encourager et à proscrire sur la base d'une analyse des pratiques existantes. L'éthique normative offre un cadre de réflexion permettant d'aider les entreprises à faire face aux enjeux associés aux DAA. Cette démarche est nécessaire pour apporter une plus grande visibilité aux données et à leur processus de traitement. L'objectif est d'anticiper les risques associés aux DAA en travaillant sur le concept de transparence. Sur le plan scientifique, les recherches sur l'éthique des données sont aujourd'hui en plein essor. Leur particularité est qu'elles adressent plusieurs disciplines qui doivent nécessairement coordonner leurs travaux afin d'offrir un cadre pertinent et clair sur les actions à mener dans ce domaine. La pluri-disciplinarité permettra ainsi d'éviter un développement cloisonné des connaissances. Ce projet s'inscrit dans cette perspective en se positionnant au carrefour de trois disciplines : le droit, l'informatique, et la gestion. A notre connaissance, il n'existe pas pour le moment d'étude similaire. L'arrivée prochaine du RGPD induit pourtant des enjeux forts qui justifient le lancement d'une recherche scientifique dédiée.

Objectifs du projet

Le projet DUTY a pour objectif d'étudier le concept de transparence dans le cadre de l'éthique des DAA dans le secteur privé, afin de proposer un programme de mise en conformité comprenant des règles normatives, un outil « ethics-by-design », et un code déontologique associé. La notion de transparence a été choisie dans la mesure où elle représente une première étape dans la réflexion éthique afin de prévenir les abus relatifs à la manipulation des DAA. Pour atteindre ce but, le projet propose d'engager une réflexion pluri-disciplinaire autour de trois Work Packages (WP) croisant chacun deux disciplines : le WP1, entre droit et gestion, a pour objectif de faire un état des lieux des pratiques d'entreprises et des normes juridiques applicables concernant l'utilisation des DAA comme un levier de performance ; le WP 2 allie droit et informatique pour déterminer le pouvoir discriminatoire des algorithmes et les cadres juridiques à développer pour réguler ces outils ; le WP 3 réunit la gestion et l'informatique pour étudier l'« ethics-by-design » : la mise en œuvre de règles éthiques lors de la conception d'outils informatiques.

Positionnement sur les scènes locale, nationale et internationale

Au niveau local, l'Université Grenoble Alpes se positionne comme un leader scientifique sur les problématiques associées aux données à travers le choix d'une approche pluri-disciplinaire. Le projet DUTY s'inscrit dans cette continuité. A l'interface de deux pôles de recherche du site, il réunit des équipes de différents laboratoires. .

Sur le plan national et européen, des textes tels que la Loi sur la République Numérique (2016), le RPDG (2018) posent un cadre renouvelé du traitement et de la gestion de données. Le projet s'inscrit dans cette dynamique de transformation en proposant un cadre à la mise en œuvre de ces réglementations, dans la perspective d'enrichir les futures lignes directrices des autorités de régulation.

Sur le plan international, le RGPD prévoit un alignement des législations étrangères aux exigences européennes par des dispositifs relevant de la « soft law » : guides de bonnes pratiques, chartes éthiques, codes déontologiques. Dans ce contexte, le projet DUTY propose d'accompagner l'extension du RGPD au-delà des frontières de l'Union Européenne.

Programme scientifique / Méthodologie / Résultats escomptés

Le projet DUTY est organisé de façon à traiter des lacunes théoriques présentes dans chaque discipline. En droit, les DAA ne sont, pour le moment, soumises à aucun cadre juridique ferme. Il est donc nécessaire de formuler des propositions afin que les règles de droit soient adaptées aux pratiques des entreprises. En gestion, les enjeux éthiques relatifs aux données sont de plus en plus reconnus. Si les pratiques d'usage et de traitement de données commencent à être étudiées, peu de recherches s'intéressent à la question de l'ethics-by-design. En informatique, rares sont les études se focalisant sur le concept de transparence. Par ailleurs, l'intégration des aspects éthiques dans le processus de conception des algorithmes se situe au stade de l'émergence et pose de nombreuses questions. Le projet DUTY vise à apporter des contributions substantielles dans chacune de ces disciplines. Des résultats conjoints sont attendus de la mise en synergie des trois disciplines afin d'apporter des réponses concrètes sur ce sujet complexe.

Le programme scientifique est structuré en Work Packages (WP) qui traitent d'un sujet à la jonction du droit, de la gestion et de l'informatique. La notion de transparence est un point commun à ces trois WP : le projet vise à obtenir une plus grande visibilité sur les pratiques des entreprises en matière de gestion des DAA (WP1), une compréhension clarifiée des risques de discrimination présents dans les algorithmes (WP 2), une inclusion de règles claires et connues lors de la conception d'outils informatiques (WP 3). Chaque WP sera principalement mis en œuvre par une équipe mixte composée de chercheurs des deux disciplines concernées. Ils travailleront avec le soutien de l'ensemble des membres du projet.

Les paragraphes suivants intègrent une description détaillée de chaque WP où sont précisés l'état actuel des connaissances à ce sujet, les objectifs, la méthodologie choisie ainsi que les résultats escomptés.

WP 1. Données anonymes et performance des entreprises, portage : CESICE et CERAG

Etat des lieux. A l'ère du numérique, l'entreprise dispose d'une quantité exponentielle de données qu'elle est tentée d'utiliser tant pour renforcer sa position sur le marché que pour optimiser sa propre performance. Une étude réalisée en 2016 par l'Autorité de la concurrence française et le Bundeskartellamt allemand présente les multiples interactions entre Big Data et libre concurrence et démontre que les données pourraient devenir le nouvel instrument de domination du marché. En témoigne l'intérêt porté par l'Autorité de la concurrence française aux implications induites par l'exploitation des données dans le secteur de la publicité en ligne (Google, Facebook...). Au sein de l'entreprise également, les données sont mobilisées pour accroître la performance : optimisation de la gestion du personnel (recrutement, paie, carrière) ou contrôle de la productivité des salariés (notamment par l'Internet des objets). Leur utilisation recèle des risques de discrimination, d'atteinte au droit à la vie privée ou à la santé des travailleurs dans la mesure où ces données anonymes sont susceptibles de repersonnalisation.

Compte tenu de la place prise par les données même anonymes dans l'économie, il apparaît urgent de mettre en perspective le droit de l'entreprise dans son ensemble, avec les droits et les libertés des personnes. Cette réflexion, plus ou moins engagée selon les disciplines, est bouleversée par l'entrée en vigueur du RGPD, obligeant les organisations à s'y conformer avant mai 2018, sous peine de se voir imposer des amendes allant jusqu'à 4 % du chiffre d'affaires annuel ou 20 millions d'euros. Malgré le niveau de protection des données personnelles nouvellement exigé, qui implique notamment la mise en œuvre de mesures techniques de sécurisation (pseudonymisation, anonymisation), le risque de repersonnalisation n'est pas cependant pas exclu en cas de circulation des données entre différents acteurs.

Objectif. L'objectif du WP 1 est double : Par l'identification des techniques d'instrumentalisation des DAA, il s'agira d'évaluer l'adaptation du cadre existant en vue de proposer une définition unitaire de la notion d'utilisation abusive des DAA, afin de garantir une plus grande loyauté dans les échanges.

Méthodologie. Pour atteindre ces objectifs, une approche en plusieurs étapes a été planifiée.

1. Revue de littérature (juridique et gestion) consistant à faire un état de l'art dans les deux disciplines : en gestion, elle consistera à identifier les dispositifs d'exploitation des données. En droit, la revue de

littérature se traduit par une étude des normes juridiques opposables aux entreprises exploitant des DAA afin d'obtenir des avantages concurrentiels ou développer leur performance interne.

2. Identification des pratiques et des dérives : collecter des données empiriques auprès d'entreprises volontaires par le biais d'entretiens, d'observations et de documents, afin d'offrir un panorama des pratiques d'entreprises à partir duquel il sera possible d'identifier les dérives potentielles.

3. Analyse des distorsions entre les pratiques actuelles et le cadre juridique existant, que la comparaison de la revue de littérature et des données empiriques permettra de révéler.

4. Elaboration d'une définition unitaire de la notion d'utilisation abusive. Cette définition sera sans doute recherchée à court terme par les régulateurs français et européens dans le cadre de la mise en œuvre du RGPD.

Résultats attendus. Les résultats attendus sont d'une part, un panorama des pratiques commerciales associées à l'utilisation des DAA par les entreprises, d'autre part, une évaluation de l'adaptation de la norme juridique au regard de ces pratiques actuelles. En conséquence, le WP 1 vise à apporter des recommandations d'ordre éthique afin de permettre une régulation à l'initiative des acteurs du marché. Un second type de contributions attendues est la formulation de normes juridiques mieux adaptées à la réalité des usages des DAA.

WP 2. Transparence des algorithmes et équité, Portage : LIG et CESICE

Etat des lieux. De nombreuses décisions sont aujourd'hui prises sur la base d'algorithmes. Or, aussi paradoxal que cela puisse paraître, les algorithmes ne sont pas neutres et la question de la discrimination algorithmique se pose aujourd'hui de façon légitime. Un algorithme est discriminatoire si les décisions qu'il prend sont différentes pour différents groupes d'individus définis par leurs motifs discriminatoires dont la liste est définie par l'article 225-1 du Code pénal. Etudier le pouvoir discriminatoire d'un algorithme consiste à déterminer quels sous-ensembles d'une population reçoivent des décisions discriminatoires. A notre connaissance, aucune étude ne propose de croiser l'étude de la discrimination algorithmique (exploration des décisions des algorithmes dans plusieurs espaces de données et définition de fonctions de comparaison) à l'étude des besoins et impacts des outils algorithmiques sur les usages afin de déterminer les cadres juridiques nécessaires à la régulation du comportement algorithmique.

Objectif. L'objectif du WP 2 est de fournir un outil informatique qui permet d'explorer le comportement algorithmique dans plusieurs cas d'usage définis en gestion, ainsi que les cadres juridiques permettant de les réguler.

Méthodologie. Notre approche sera incrémentale et basée sur l'interaction agile entre les disciplines. Elle s'appuie sur les étapes suivantes :

1. Définition des besoins et usages en gestion : Ces cas d'usage seront notre point de départ et permettront de cadrer les décisions algorithmiques auxquelles nous nous intéresserons dans ce travail.

2. Identification d'un cadre juridique de la discrimination algorithmique pour les cas d'usage : Cette étape s'effectuera sur la base des résultats d'un projet CNRS MASTODONS ELM (Employees Lives Matter) réalisé en 2017 concernant la discrimination dans les plateformes participatives (cf annexe 2).

3. Développement de l'outil informatique : Nous avons aujourd'hui outil pour tester algorithmiquement différentes distances entre des décisions d'assignation de tâches sur les plateformes participatives (projet ELM décrit en annexe 2). Nous proposons de l'étendre avec un simulateur pour étudier la discrimination algorithmique dans les cas d'usage définis en 1.

Résultats attendus. Le travail sur le WP permettra d'offrir un outil informatique pour explorer différents scénarios d'usage et déterminer les algorithmes les plus pertinents dans la prise de décision. L'outil constituera un environnement de test des prises de décision algorithmiques pour les gestionnaires et les

juristes. Cet outil apportera donc la transparence qui permettra de mieux comprendre les risques et potentiels de la décision algorithmique.

WP 3. Responsabilisation des acteurs par l'ethics by design. Portage : CERAG et LIG

Etat des lieux. Le développement d'une société de l'information porté par des outils technologiques de plus en plus sophistiqués induit des enjeux éthiques importants générés par les risques d'atteinte aux personnes. Ainsi, les questionnements concernant les notions de vie privée, de propriété intellectuelle, d'accessibilité à l'information ont désormais place dans le débat public. Pour interroger et tenter de répondre à ces enjeux, la littérature scientifique s'est principalement centrée sur la régulation des usages. Ces approches ont pour point commun d'intervenir une fois que l'outil informatique est développé et prêt à être utilisé, mais n'abordent pas le processus décisionnel qui sous-tend la conception des algorithmes. Les études scientifiques sur ce thème se développent à peine. Le WP 3 se positionne dans ce domaine de recherche émergent, à la fois en informatique et en gestion. Il rejoint la dimension appliquée de l'éthique des données qui suppose la réalisation de cas d'usage conduisant à l'intégration de règles éthiques dès la conception des outils. Cette approche se positionne en complément d'une action sur les comportements visant à une plus grande responsabilisation des acteurs.

Objectif. L'objectif du WP 3 est d'étudier le développement d'une approche « ethic-by-design » dans le cadre du développement d'outils informatiques traitant des DAA. Le WP 3 complète le travail de réflexion sur les pratiques et les usages présentés dans les WP 1 et 2 en rendant concrète la notion de transparence abordant celle-ci à l'échelle de l'outil.

Méthodologie. La méthodologie choisie est qualitative en raison du caractère exploratoire de l'étude proposée. Après un premier état de l'art, deux études de cas seront menées et permettront de suivre au plus près le développement d'un outil technique et d'étudier comment les règles et normes sont intégrées au cours de la conception. Trois étapes définissent cette approche :

1. Etat de l'art autour du concept d'éthique des données, définition du cadre théorique.
2. Etude de cas n°1 : les données sont hébergées sur une plateforme appartenant au département RH d'une entreprise et un algorithme est utilisé pour déterminer l'attribution des primes.
3. Etude de cas n°2 : les données sont hébergées sur une plateforme appartenant au département contrats d'une entreprise et un algorithme est utilisé pour comparer les réponses à un appel d'offres.

Ces deux cas concrets impliquent la conception d'un outil informatique adapté intégrant des règles permettant de protéger les données en entrée à partir d'algorithmes de décision qui, dans les deux cas, doivent être transparents.

Résultats attendus. Notre étude a vocation à fournir une compréhension approfondie des mécanismes par lesquels il est possible de prévenir ou de réduire les pratiques déloyales en intégrant dès la conception de l'outil des règles éthiques clairement définies. Un deuxième résultat attendu est un guide de bonnes pratiques à l'usage des développeurs où transparaissent clairement des règles éthiques conduisant à la transparence des algorithmes qui sous-tendent le processus décisionnel. A partir des cas étudiés, nous prévoyons d'intégrer les pratiques préconisées par ce guide aux outils de gestion de projet informatique afin de soutenir le développement d'un programme de mise en conformité nécessaire aux entreprises.

Références

- Luca, M., and Fisman, R. Fixing discrimination in online marketplaces. Harvard Business Review 2016.
- Tramer, F.; Atlidakis, V.; Geambasu, R.; Hsu, D. J.; Hubaux, J.; Humbert, M.; Juels, A.; and Lin, H. Discovering unwarranted associations in data-driven applications with the fairest testing toolkit. CoRR abs/1510.02377, 2015.
- Smith, H. J., & Hasnas, J. (1999). Ethics and Information Systems: The Corporate Domain. MIS Quarterly, 23(1), 109-127.
- Banerjee, D., Cronan, T. P., & Jones, T. W. (1998). Modeling IT Ethics: A Study in Situational Ethics. MIS Quarterly, 22(1), 31-60.

Organisation du projet : échéancier, personnel impliqué et pourcentage du temps prévu, partenariats, utilisation des crédits

Echéancier global du projet

		2018				2019				2020
		T2	T3	T4	T1	T2	T3	T4	T1	
WP1	WP1.1 Revue de littérature									
	WP1.2 Définition de l'utilisation abusive									
	WP1.3 Identification des pratiques et des dérives									
	WP1.4 Analyse des distorsions									
WP2	WP2.1 Définition des besoins									
	WP2.2 Identification du cadre juridique									
	WP2.3 Développement de l'outil									
WP3	WP3.1 Etat de l'art et cadre théorique									
	WP3.2 Etude de cas n°1 (RH)									
	WP3.3. Etude de cas n°2 (Contrats)									

Personnel impliqué et pourcentage de temps prévu

Emilie Hoareau (CERAG, Maître de conférences) : 40% du temps recherche
 Karine Samuel (CERAG, Professeur des Universités) : 20% du temps recherche
 Sihem Amer-Yahia (CNRS, Directrice de Recherche) : 30 % du temps recherche
 Oana Goga (CNRS, Chargée de recherche) : 15 % du temps recherche
 Aïda Bennini (CESICE, Maître de conférences) : 25% du temps recherche
 Sabrina Mraouahi (CRJ, Maître de conférences) : 15% du temps recherche
 3 stagiaires : 100% du temps recherche

Partenariats

Prof Shady el Bassuoni, American Univeristy de Beirouth (discrimination)
 Pr. Dr. Hendrik Send, Alexander von Humboldt-Institut für Internet und Gesellschaft GmbH, Berlin
 Centre de recherche de droit des affaires et du commerce international de l'Université de Montréal (UDM)
 Association Transeurope Expert (collectifs d'enseignant-chercheurs en droit privé)

Utilisation des crédits

Dépenses de fonctionnement	
Nombre de mois de stage : Rémunération de stagiaires (3 stages de 6 mois) : 1 stagiaire par WP	10.080€
Frais de mission (dont visiteurs) : Participation à des conférences académiques nationales et internationales (3 conférences à l'international pour 2 personnes) :	
- Colloque ICIS, San Francisco (4 jours pour 2 personnes)	4.000€
- Mission à Beyrouth ou invitation du partenaire (5 jours)	4.000€
- Visite de recherche au centre du droit des affaires et du commerce international pour partenariat (HEC Montréal) (5 jours)	4.000€
- Missions à Paris (CNIL et autorité de la concurrence)	4.000€
Actions de communication : Organisation de workshops à l'UGA sur la gouvernance économique des données	1.500 €
Equipement : Petit matériel/Consommables : achat de 2 ordinateurs	1.920 €
Autres dépenses : documentation, abonnements, ouvrages	500 €
TOTAL	30.000 €