

Compte-rendu de la journée « Données personnelles, vie privée et éthique (EGC&IA 2018) »

*Organisée par l'AFIA et EGC
le mercredi 23 Janvier 2018
à la Maison des sciences et de l'Homme de Paris Nord*

Responsables scientifiques

- Jérôme Azé (Université de Montpellier/LIRMM)
- Thierry Charnois (Université Paris 13/LIPN)

Comité de programme

- Maxime Amblard, LORIA
- Jean-Yves Antoine, Université de Tours
- Annie Blandin, IMT Atlantique
- Gauvain Bourgne, CNRS & Sorbonnes Universités, UPMC Paris 06, LIP6
- Nora Cuppens-Bouahia, IMT Atlantique
- Gaël Dechalendar, CEA
- Karën Fort, Paris Sorbonne
- Philippe Lenca, IMT Atlantique
- Fabrice Muhlenbach, Université Jean Monnet
- Guillaume Piolle, Centrale Supélec

Page Web

- <https://egc-ia-2018.sciencesconf.org/>

La 3e journée thématique EGC & IA : Données personnelles, vie privée et éthique (EGC&IA 2018) s'est tenue le 23 Janvier 2018 à Paris, à la Maison des sciences et de l'Homme de Paris Nord. Cette édition visait à offrir un panorama des approches liées à la collecte, le stockage et la gestion des données personnelles en lien étroit avec les problématiques de vie privée et d'éthique.

Thématiques de la journée

À l'ère de Big Data, chaque utilisateur de notre monde hyper-connecté laisse derrière lui une myriade de traces numériques tout en effectuant ses activités quotidiennes. Ces masses de données sont majoritairement exploitées par de grandes compagnies afin d'en extraire de précieuses informations permettant de modéliser le comportement humain et d'améliorer des stratégies de marketing. Chaque compagnie n'a qu'une vision limitée des données personnelles disponibles, uniquement celles collectées via les services qu'elle propose.

Et surtout, chaque utilisateur a un contrôle assez limité sur les données réellement collectées, et sur l'usage qu'il pourrait en faire. Des solutions telles que les "Personal Data Store" ou les "Information Management System" apparaissent.

L'objectif est de permettre aux utilisateurs de centraliser des données fréquemment utilisées par des sites marchands ou autres et ainsi de faciliter les actions de l'utilisateur lorsqu'il a besoin de transmettre ces informations en lui évitant de devoir les ressaisir (par exemple).

Il y a cependant un manque réel d'algorithmes et de modèles dédiés pour le traitement des données personnelles afin d'y détecter des comportements et d'en extraire de la connaissance,

tout en garantissant les aspects liés à la protection de la vie privée et les aspects liés à l'éthique dans le cas de scénario d'utilisation centré sur l'utilisateur.

Aujourd'hui, l'analyse des données personnelles, la protection individuelle de la vie privée et le respect de l'éthique sont les éléments clés pour tirer parti des services pour un nouveau type de systèmes.

La disponibilité d'outils d'analyse personnelle capables d'extraire des connaissances cachées à partir de données individuelles tout en protégeant le droit à la vie privée et en respectant l'éthique peut aider la société à passer de systèmes centrés sur l'organisation à des systèmes centrés sur l'utilisateur, où l'utilisateur est le propriétaire de ses données personnelles et peut gérer, comprendre, exploiter, contrôler et partager ses propres données et les connaissances pouvant en être extraites de manière complètement sûre.

Cette problématique trouve un écho particulier dans le cadre du nouveau Règlement européen sur la Protection des Données Personnelles (RPDG) paru au journal officiel de l'Union européenne et entrant en application le 25 mai 2018. Cette réglementation renforce les droits des personnes, notamment par la création d'un droit à la portabilité des données personnelles, et responsabilise les acteurs traitant des données (responsables de traitement et sous-traitants). (<https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-ce-qui-change-pour-les-professionnels>)

Programme

14h – 15h : Présentations de l'AFIA, par Yves Demazeau (Président de l'AFIA) et de EGC par Fabrice Guillet (Président de EGC)

15h – 16h30 : Session Données personnelles, vie privée et éthique

15h – 15h30 : « A Declarative Approach to Linked Data Anonymization », Angela Bonifati, Rémy Delanaux, Marie-Christine Rousset and Romuald Thion.

15h30 – 16h : « Co-clustering for differentially private synthetic data generation », Tarek Benkhelif, Françoise Fessant, Fabrice Clérot and Guillaume Raschia.

16h – 16h30 : « Système de gestion de l'obsolescence dans une base d'informations personnelles », Salma Chaieb, Véronique Delcroix, Ali Benmrad and Emmanuelle Grislin.

16h30 – 17h : Pause

17h – 18h : Conférence de Karën Fort (Université Paris-Sorbonne)
"TAL et éthique"

Cette journée était organisée par Jérôme Azé (Université de Montpellier/LIRMM) et Thierry Charnois (Université Paris 13/LIPN). Elle a été ouverte par Yves Demazeau (CNRS, Président de l'AFIA) qui a présenté les activités et perspectives de l'AFIA et Fabrice Guillet (Université de Nantes/Polytech'Nantes, Président de EGC) qui a présenté les activités de l'association EGC.

Suite au faible nombre de soumissions, la journée initialement prévue a été réduite à une demi-journée. Après la présentation des associations AFIA et EGC, trois orateurs se sont succédés, en ciblant différents champs de la discipline :

« A Declarative Approach to Linked Data Anonymization », Angela Bonifati, **Rémy Delanaux**, Marie-Christine Rousset and Romuald Thion.

Résumé : We present a declarative approach for privacy-preserving data publishing in Linked Open Data, in which privacy and utility policies along with basic data transformation operators are encoded as SPARQL queries. We believe that our approach can be useful in many settings in which data providers are responsible for the non-disclosure of information that could serve as quasi-identifiers whenever crossed with external data sources. Data producers need their own means to specify the privacy policies they want to enforce on their datasets to be published, but also to specify the utility policies that tell which parts of the data should be kept. They also need tools for computing the anonymization operations to apply to their datasets prior to their publication with the guarantee that the specified privacy and utility policies are verified by the resulting dataset. Our approach is designed to meet these requirements by leveraging the expressive power of SPARQL queries and the effectiveness of SPARQL query engines.

« Co-clustering for differentially private synthetic data generation », **Tarek Benkhelif**, Françoise Fessant, Fabrice Clérot and Guillaume Raschia.

Résumé : We propose a methodology to anonymize microdata (i.e. a table of n individuals described by d attributes). The goal is to be able to release an anonymized data table built from the original data while meeting the differential privacy requirements. The proposed solution combines co-clustering with synthetic data generation to produce anonymized data. First, a data independent partitioning on the domains is used to generate a perturbed multidimensional histogram; a multidimensional co-clustering is then performed on the noisy histogram resulting in a partitioning scheme. This differentially private co-clustering phase aims to form attribute values clusters and thus, limits the impact of the noise addition in the second phase. Finally, the obtained scheme is used to partition the original data in a differentially private fashion. Synthetic individuals can then be drawn from the partitions. We show through experiments that our solution outperforms existing approaches and we demonstrate that the produced synthetic data preserve sufficient information and can be used for several datamining tasks.

« Système de gestion de l'obsolescence dans une base d'informations personnelles », Salma Chaieb, **Véronique Delcroix**, Ali Benmrad and Emmanuelle Grislin.

Résumé : L'évaluation régulière du risque de chute des personnes âgées requiert des informations fiables et nombreuses. Comme il n'est pas possible de recueillir régulièrement toutes ces informations, les observations sont faites au fil du temps et conservées, ce qui entraîne une problématique liée au vieillissement des informations. Cet article traite de la détection des informations obsolètes dans une base d'informations sur une personne âgée. Nous proposons un agent chargé de maintenir une base d'informations et capable de fournir à la demande des informations fiables et cohérentes sur la personne. L'agent est équipé d'un modèle de connaissances sur les personnes âgées, sous forme d'un réseau bayésien et d'un module de raisonnement chargé de la détection et de la gestion des contradictions et des doutes sur les informations.

Puis, après cette session de présentation d'articles, Karën Fort (Université Paris-Sorbonne) a donné une conférence d'une heure sur la thématique « TAL et éthique ».

Karën Fort a dans un premier temps présenté les spécificités des problématiques éthiques en TAL (Traitement Automatique des Langues), puis elle a élargit la problématique aux enjeux éthiques de la communauté scientifiques. Elle a ensuite présenté les actions du groupe « Ethique et TAL ». Elle a conclu en montrant la nécessité d'une réflexion éthique à différents niveaux : chercheurs, sociétales, politique, ...