

# Un logiciel d’anonymisation ARX à comprendre, à étendre.

Jean-François COUCHOT

## 1 Présentation

ARX<sup>1</sup> est un logiciel open source ayant pour objectif d’anonymiser des données personnelles sensibles.

Nombreux sont les modèles de protection de la vie privée et leurs attaques associées à être pris en compte par l’outil. Par exemple, le  $k$ -anonymat [Swe02], la  $l$ -diversité [MKG07] sont autant de modèles qui peuvent être vérifiées par l’outil.

Nombreuses sont les méthodes de transformation des données à être implantées dans l’outil : généralisation, catégorisation, . . . , certains algorithmes locaux [EPK14] ou globaux de confidentialité différentielle [DMNS06].

ARX implante des méthodes permettant d’analyser l’utilité des données : les généralisations, par exemple, induisent des pertes de précision et donc une réduction de l’utilité des données.

Enfin ARX autorise l’intégration de nouveaux plugins, permettant ainsi d’en étendre les fonctionnalités et d’en faire bénéficier ensuite la communauté.

## 2 Objectifs

1. Déployer ARX et reproduire, en documentant, les nombreux scénarios d’utilisation détaillés à la page de publications<sup>2</sup>. Le choix des scénarios retenus fait partie du travail demandé.
2. Reproduire dans le contexte d’ARX les expérimentations d’anonymisation obtenues dans l’article [CGR19] basées sur des méthodes de transformation usuelles.
3. Déployer dans ARX un plugin pour GRR [WBLJ17], la publication de données multidimensionnelles réduites [WXY<sup>+</sup>19], une extension non encore publiée et la dissociation associée à son modèle inverse [ACaBP20].

## Références

- [ACaBP20] Nancy Awad, Jean-François Couchot, Bechara al Bouna, and Laurent Philippe. Publishing anonymized set-valued data via disassociation towards analysis. *Future Internet*, 12(4) :71, 2020.
- [CGR19] Jean-François Couchot, Christophe Guyeux, and Guillaume Royer. Anonymously forecasting the number and nature of firefighting operations. In Bipin C. Desai, Dimosthenis Anagnostopoulos, Yannis Manolopoulos, and Mara Nikolaidou, editors, *Proceedings of the 23rd International Database Applications & Engineering Symposium, IDEAS 2019, Athens, Greece, June 10-12, 2019*, pages 30 :1–30 :8. ACM, 2019.
- [DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.
- [EPK14] Úlfar Erlingsson, Vasily Pihur, and Aleksandra Korolova. Rappor : Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.
- [MKG07] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramanian.  $l$ -diversity : Privacy beyond  $k$ -anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1) :3–es, 2007.
- [Swe02] Latanya Sweeney.  $k$ -anonymity : A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05) :557–570, 2002.
- [WBLJ17] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*, pages 729–745, 2017.
- [WXY<sup>+</sup>19] Ning Wang, Xiaokui Xiao, Yin Yang, Jun Zhao, Siu Cheung Hui, Hyejin Shin, Junbum Shin, and Ge Yu. Collecting and analyzing multidimensional data with local differential privacy. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 638–649. IEEE, 2019.

---

1. <https://arx.deidentifier.org/>

2. <https://arx.deidentifier.org/publications/>