**ANLYSE DE LA BASE PRC (Privacy right clearinghouse)**

C’est une base aux Etats Unis qui rassemble les notifications de violations des données signalées publiquement par les agences gouvernementales américaines dans une base de données unique et consultable.

Cette base contient environ : 76 987 notifications de violation de données, 27 833 lettres de notification, 8.28 milliards de personnes touchées.

Variable dans la base :

* Type d’organisation : BSF, MED, …etc.
* Méthode de violation : Hack, Port, … etc.
* Classification UNKN (inconnu)

Variable intéressante :

* Id
* organization\_type
* incident\_details
* breach\_type
* reported\_date
* breach\_date
* end\_breach\_date
* total\_affected
* residents\_affected
* breach\_location\_city
* breach\_location\_state
* breach\_location\_country

PRC database does not directly provide the loss associated with an event, but reports the number of records (that is the number of user accounts) affected by the breach.

Bonjour Madame, Monsieur,

Je me permets de vous écrire car j’ai besoin de la base données Privacy Right Clearinghouse (PRC) dans le cadre de mes travaux de recherche de fin d’étude académique.

En effet, je suis étudient au cycle ingénieur à l’ENSAE Paris, France, où je poursuis mes études en Actuariat. Dans le cadre de la rédaction de mon mémoire de fin d’étude, je travaille sur la modélisation de la sévérité du risque cyber en assurance.

Compte tenu du manque de données criarde pour réaliser cette étude, la base Privacy Right Clearinghouse (PRC) se veut être une ressource incontournable pour mener à bout ce projet.

D’ailleurs, je travaille sous la supervision de Hugo RAPIOR, l’un des auteurs de l’article [CYBER INCIDENT REPORTS : extrapolating severity using neural networks](https://detralytics.com/wp-content/uploads/2023/05/Detra-Note-2023-4_Cyber-Incident-Reports.pdf) qui est une étude analysant des données textuelles de notification de violation de données pour prédire la sévérité d’une attaque cyber en utilisant les réseaux de neurone. Dans cet article la base de données utilisées est la base PRC.

Nous voulons dans le cadre de notre travail approfondir les travaux de cet article en utilisant de l’IA générative et les techniques de Machine Learning beaucoup plus poussée (NLP, LSTM, CNN, etc..) Pour modéliser la sévérité du risque ; si possible voir comment ou peut évaluer un Solvency Capital Requirement (SCR) associer au risque cyber pour les compagnies d’assurance.

En espérant une réponse favorable, je vous prie d’agréer Madame, Monsieur, l’expression de mes sentiments les plus distinguées.

Dear Privacy Rights Clearinghouse Team,

I am writing to respectfully request access to the Privacy Rights Clearinghouse (PRC) data breach database for academic research purposes.

My name is Dountio Zaboue Franck Osée, and I am a final-year engineering student at ENSAE Paris (France), specializing in actuarial science and risk modeling. Under the supervision of Hugo Rapior, co-author of the paper *“*[*Cyber Incident Reports: Extrapolating Severity Using Neural Networks*](https://detralytics.com/wp-content/uploads/2023/05/Detra-Note-2023-4_Cyber-Incident-Reports.pdf)*”*, I am currently conducting my master’s thesis on the modeling of cyber risk severity in the insurance sector.

Building upon previous work that used your PRC dataset, my research aims to apply advanced machine learning and generative AI techniques, including LSTM, CNN, and NLP models to predict the severity of cyber incidents based on textual breach reports. A long-term objective of the project is to explore how these models can support the estimation of Solvency Capital Requirements (SCR) for cyber risks in regulatory and actuarial contexts.

This research directly supports your mission to enhance public awareness of privacy issues and improve consumer privacy protections. By helping insurers and policymakers better understand the consequences of data breaches, our work contributes to more resilient responses to digital privacy threats.

As an academic researcher working on a non-commercial project, I would be very grateful to receive access to the PRC dataset under your complimentary academic access tier.

Thank you for considering my request. Please let me know if any additional information or documentation is required.

Warm regards,

Dountio Zaboue Franck Osée  
Engineering Student - Actuarial Science  
ENSAE Paris  
franckosee.dountiozaboue@ensae.fr