

## ÉTUDE DE CAS



## L'ICIJ

## Comment l'ICIJ s'est-elle servie de Neo4j pour révéler l'affaire Swiss Leaks

**SECTEUR**

Médias

**OBJECTIFS**

Aider les journalistes à exploiter les données complexes de l'affaire Swiss Leaks pour approfondir leurs enquêtes

**STRATÉGIE**

- Mettre en place un nouvel outil d'analyse de bases de données de graphe pour permettre aux journalistes de découvrir et d'explorer rapidement et facilement de nouveaux faits pour leurs articles
- Partager l'outil dans le monde entier pour favoriser la collaboration autour du projet Swiss Leaks

**SOLUTION**

- Neo4j s'est imposé comme l'outil idéal pour la gestion de données complexes, de recherches et de requêtes
- L'outil de visualisation de Linkurious a mis le processus d'analyse à la disposition de tous les journalistes de l'ICIJ

**RÉSULTATS**

- Permet aux journalistes du monde entier de découvrir et de faire des recherches rapidement et facilement dans des données complexes et non structurées
- La bases de données de graphe révèle des faits et interactions exceptionnels qu'il était impossible de mettre à jour auparavant
- L'outil a changé les habitudes en termes d'analyse des fraudes et transformé le processus du journalisme d'investigation

Démontrer l'efficacité du graphe dans le journalisme d'enquête.

Au début de l'année 2014, deux journalistes français du Monde, Gérard Davet et Fabrice Lhomme, ont eu accès à un jeu de données très complexe mais néanmoins d'une grande valeur. Ces données concernaient les comptes bancaires de plus de 100 000 clients de la banque HSBC en Suisse, mettant ainsi au jour de nombreuses pratiques contraires à l'éthique. Davet et Lhomme y ont vu la possibilité de rédiger un article détonnant sur la fraude, l'évasion fiscale et le crime international. Le problème ? Les données étaient trop complexes pour être analysées selon des méthodes conventionnelles.

### L'objectif : Venir à bout de la complexité des données pour révéler des informations précieuses

En règle générale, les journalistes doivent tenter d'identifier des relations entre les données dans des fichiers Excel, mener des recherches manuelles sur Internet et parfois établir physiquement des liens entre les personnes et les entités pour obtenir des faits intéressants pour leurs articles.

Toutefois, Davet et Lhomme ont reconnu que le jeu de données Swiss Leaks était trop complexe pour être analysé manuellement ou par eux seuls. Ils se sont donc tournés vers le Consortium International des Journalistes d'Investigation (International Consortium of Investigative Journalists, ICIJ), lançant ainsi la collaboration journalistique la plus importante au monde.

Mar Cabra, éditrice du département des données et des recherches (Data and Research Unit) chez l'ICIJ, savait qu'ils auraient besoin d'un outil pour analyser de manière plus efficace les relations entre les données pour les enquêtes actuelles et futures.

Les données Swiss Leaks contenaient des informations sur des propriétaires de comptes HSBC situés dans plus de 200 pays et d'une valeur totale de plus de 100 milliards de dollars. Mais ces informations étaient réparties dans des milliers de fichiers sans liens apparents les uns avec les autres. En raison de la complexité des données, Cabra et l'ICIJ avaient besoin d'un outil pour analyser cette immense quantité de données non structurées et pour les comprendre rapidement et aisément.

D'autre part, Mar Cabra ne voulait pas que les spécialistes et les développeurs de données deviennent des intermédiaires. Cabra souhaitait que le processus de découverte et d'analyse des données soit directement accessible aux journalistes du monde entier, quelle que soit leur expertise technique.

### La solution : « Suivre l'argent à la trace » grâce à l'analyse de graphes

Avant l'affaire Swiss Leaks, l'organisation avait déjà été confrontée à des jeux de données complexes. Ainsi, Cabra savait que l'ICIJ aurait besoin d'une solution de base de données de graphe.

« En travaillant sur des affaires telles que l'Offshore Leaks, j'ai compris l'importance du rôle de l'analyse de graphes dans les enquêtes sur la corruption financière », explique Cabra. « Les connexions sont essentielles pour saisir le nœud de l'intrigue : elles vous montrent qui fait

## ÉTUDE DE CAS



« En utilisant Neo4j et la technologie Linkurious, nous avons complètement changé la donne. »

- Mar Cabra

affaire avec qui. Nous avons décidé très tôt qu'il nous faudrait utiliser une approche orientée graphe pour l'affaire HSBC. »

Le département de données et de recherches a commencé par recréer la base de données des clients HSBC à partir des fichiers Excel. Ensuite, il a associé chaque nom à un ou plusieurs pays (chacun de ces éléments étant appelés « nœuds » dans la base de données de graphe). Enfin, il a converti les données en un format graphique pour étudier les connexions entre les nœuds.

Au total, les données représentaient quelque 60 000 fichiers contenant des informations sur plus de 100 000 clients dans 203 pays. La base de données de graphe obtenue comptait plus de 275 000 nœuds et 400 000 relations entre eux.

L'ICIJ a travaillé aux côtés de Talend, une société spécialisée dans les logiciels d'intégration open source, pour transférer le jeu de données d'origine dans la base de données de graphe Neo4j de Neo Technology. Un autre partenaire de Neo, Linkurious, a fourni une application Web servant d'interface utilisateur de sorte que la base de données de graphe soit facilement accessible aux journalistes.

L'approche de visualisation des graphes a permis aux journalistes de l'ICIJ d'identifier les connexions entre les personnes et les comptes bancaires, leur donnant ainsi la possibilité de « pister l'argent » pour identifier des dizaines de cas de fraude, de corruption et d'évasion fiscale.

Voici un exemple illustrant ces types de relations dans un graphe.

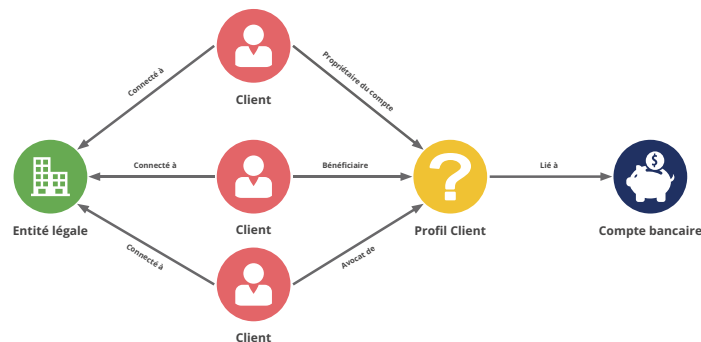


Schéma représentant les entités et les connexions dans le jeu de données Swiss Leaks.

### Les résultats : Exploiter des données complexes pour enquêter sur la fraude

Après avoir importé les données dans Neo4j et Linkurious, Cabra et son équipe ont remarqué une différence sensible dans l'aptitude des journalistes à exploiter les données Swiss Leaks.

« En utilisant Neo4j et la technologie de Linkurious, nous avons complètement changé la donne », explique Cabra. « Ces outils ont permis à nos journalistes d'explorer des données financières complexes selon des méthodes très différentes de celles que nous utilisions auparavant. »

La solution Neo4j-Linkurious a même répondu à la demande de Cabra en termes de simplicité d'utilisation.

« Elle était tellement intuitive que les journalistes n'ont eu besoin que d'une formation très courte », affirme Cabra. « Nous avons simplement mis en ligne une démonstration rapide et un webinaire. Nos journalistes ont été surpris de la facilité avec laquelle il était possible de faire une requête sur les données : ils n'ont pas eu besoin de compétences informatiques avancées pour le faire. »

Une fois l'outil partagé dans la salle de presse virtuelle de l'ICIJ, des journalistes du monde entier ont ainsi exploité les données et l'outil d'analyse de graphes depuis leurs zones géographiques respectives, envoyant des requêtes à l'échelle mondiale. Grâce à la visualisation simple des réseaux entourant les clients et les comptes, ils ont découvert bien

## ÉTUDE DE CAS

“Il s’agit d’un outil de découverte révolutionnaire qui a transformé notre processus de journalisme d’investigation.”

– Mar Cabra

plus de connexions qu'auparavant et ont pu révéler de nouvelles affaires qui, par la suite, ont fait la une des journaux du monde entier. Jusqu'à présent, les journalistes isolés établissaient des connexions à la main à partir d'informations issues de dizaines de fichiers : une tâche chronophage aux résultats parfois inexacts.

En utilisant Neo4j pour étudier les données divulguées de HSBC, les journalistes ont pu facilement identifier les acteurs principaux, les intermédiaires et les bénéficiaires liés au scandale (quelle que soit leur localisation) et ont pu constater en un clic les liens existants entre eux.

L'outil de visualisation Linkurious étant compréhensible à travers les frontières culturelles et géographiques, il a constitué un moyen idéal de mettre en place une collaboration internationale autour du projet, et s'est révélé un outil inestimable aussi bien pour les reportages que pour la vérification des faits lors de la rédaction des articles par les journalistes.

« Il s'agit d'un outil de découverte révolutionnaire qui a transformé notre processus de journalisme d'investigation », fait remarquer Cabra. « Nous disposons d'un immense potentiel pour utiliser les graphes dans le but de découvrir d'autres scandales financiers ou d'entreprise. Nous n'aurions jamais pu réussir à cette échelle, c'est assez exceptionnel ! »

Aujourd'hui, Neo4j est devenu partie intégrante de l'ICIJ dans le cadre de ses projets liés au big data. Suite aux révélations de l'affaire Swiss Leaks, certains journalistes externes se sont également adressés à l'ICIJ pour utiliser les outils d'analyse de graphe.

### Les conséquences : Exploiter des données complexes pour enquêter sur la fraude

En février 2015, plus de 50 agences de presse internationales (dont Le Monde) ont révélé comment HSBC est venue en aide à des criminels, des trafiquants et des évadés fiscaux et a fait affaire avec eux en protégeant plus de 100 000 clients dont les comptes représentent plus de 100 milliards de dollars en Suisse.

Le projet de six mois, auquel 150 journalistes ont participé, a été récompensé dans la catégorie Investigation de l'année des Data Journalism Awards organisés par le Global Editors Network. Aujourd'hui, la révélation des affaires se poursuit.

### À propos de l'ICIJ : Les meilleurs enquêteurs du quatrième pouvoir

L'International Consortium of Investigative Journalists (ICIJ) est un réseau global de plus de 190 journalistes d'investigation exerçant dans plus de 65 pays et collaborant sur des reportages d'investigation détaillés.

Fondée en 1997 par Chuck Lewis, un journaliste américain, l'ICIJ a été lancée dans le cadre d'un projet du Centre pour l'Intégrité Publique (Center for Public Integrity) afin de développer le journalisme de type « chien de garde » du centre, se concentrant ainsi sur des sujets tels que le crime international, la corruption et la responsabilité du pouvoir. Soutenus par le centre et ses spécialistes en reportage assisté par ordinateur, ses experts en dossiers publics, ses fact-checkers (vérification des faits) et ses avocats, les reporters et éditeurs de l'ICIJ fournissent aux journalistes du monde entier des ressources en temps réel ainsi que les outils et techniques les plus récents.

#### À propos de Neo Technology

Neo Technology est le créateur de Neo4j, leader mondial des bases de données de graphe, qui propulse les relations entre les données au premier plan. Depuis la recommandation de produits et de services personnalisés par les entreprises, aux sites qui offrent des fonctionnalités de réseaux sociaux, en passant par le diagnostic des problèmes réseau par les opérateurs de services de télécommunications, la réinvention des modèles d'accès, d'identité et de données maître, les entreprises adoptent les bases de données de graphe pour modéliser, stocker et interroger aussi bien les données que les relations existant entre elles. De grandes entreprises telles que Walmart, eBay, UBS, Nomura, Cisco, HP et Telenor, ainsi que des start-ups telles que CrunchBase, Medium, Polyvore et Zephyr Health utilisent Neo4j pour tirer parti des relations entre les données.

**UK** [uk@neotechnology.com](mailto:uk@neotechnology.com)  
**France** [ventes@neotechnology.com](mailto:ventes@neotechnology.com)  
**Nordics** [nordics@neotechnology.com](mailto:nordics@neotechnology.com)  
**DACH** [vertrieb@neotechnology.com](mailto:vertrieb@neotechnology.com)  
**Southern Europe** [southern-europe@neotechnology.com](mailto:southern-europe@neotechnology.com)