

## CASE STUDY



## The ICIJ

## How the ICIJ Used Graph Database Technology to Uncover the Swiss Leaks Story

## INDUSTRY

Media

## CHALLENGE

Help reporters break through the complex Swiss Leaks data for better investigative journalism

## STRATEGY

- Set up a new graph data analysis tool to enable journalists to explore and discover new facts for their stories, quickly and easily
- Share the tool worldwide to allow collaboration on the Swiss Leaks project

## SOLUTION

- Neo4j was the perfect tool for managing complex data, searches and queries
- Linkurious visualisation tool made the analysis process accessible to all ICIJ reporters

## RESULT

- Allows journalists worldwide to search and discover complex, unstructured data quickly and easily
- Graph database uncovers ground-breaking angles and stories that weren't possible to spot before
- A game-changer for fraud analysis transforming the investigative journalism process

Demonstrating the power of the graph in investigative journalism.

In the beginning of 2014, two French journalists from Le Monde, Gérard Davet and Fabrice Lhomme, received access to a massively complex – yet highly valuable – dataset. The data detailed over 100,000 clients and their bank accounts at the Swiss branch of HSBC, casting light on various unethical practices.

Davet and Lhomme saw the data's potential for a ground-breaking story on fraud, tax evasion and international crime. The problem? The data was simply too complex to be analysed by conventional means.

### The Challenge: Breaking through Data Complexity to Valuable Insights

Traditionally, reporters have to try and spot relationships between data in Excel files, conduct manual Internet searches and sometimes physically draw out connections between people and entities to get the right facts for their stories.

However, Davet and Lhomme recognized that the Swiss Leaks dataset was simply too complex to analyse manually or by themselves. So they turned to the International Consortium of Investigative Journalists (ICIJ), which started one of the biggest journalistic collaborations of all time.

Mar Cabra, editor of the Data and Research Unit at the ICIJ, knew they would need a tool that could better analyse the relationships in the data for both this and future investigations.

The Swiss Leaks data included information from HSBC account holders located in more than 200 countries collectively holding sums in excess of \$100 billion. But their information was scattered in thousands of files with no straightforward connection among each other. The complexity of the data meant Cabra and the ICIJ needed a means of analysing the vast amounts of unstructured data and making sense of it quickly and easily.

Cabra also had one more requirement: She didn't want data scientists or developers to have to serve as intermediaries. Cabra wanted the data discovery and analysis process to be accessible to reporters worldwide – no matter what their technical expertise.

### The Solution: Graph Analysis Holds the Key to 'Follow the Money'

The Swiss Leaks story wasn't the organisation's first experience with a complex dataset. As a result, Cabra knew the ICIJ would need a graph database solution.

"While working on stories like Offshore Leaks, I learned how important graph analysis is when investigating financial corruption," Cabra said. "Connections are key to under-

## CASE STUDY



"Using Neo4j and Linkurious technology was a complete game-changer for us."

– Mar Cabra

standing what the real story is: they show you who's doing business with whom. We decided early on that we needed to use a graph-based approach for the HSBC Leaks."

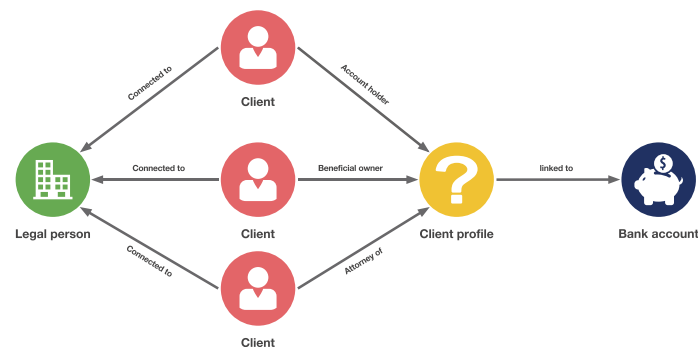
The Data and Research Unit's first move was to recreate the HSBC client database from the provided plain Excel files. Next, they connected every name to one or several countries (both referred to as the 'nodes' in the graph database). Finally, they turned the data into a graph format to explore the connections between nodes.

In total, the leak held around 60,000 files that contained information about over 100,000 clients in 203 countries. The resulting graph database had more than 275,000 nodes with 400,000 relationships among them.

The ICIJ worked with open source integration software specialist Talend to transfer the original dataset into Neo Technology's Neo4j graph database. Another Neo partner, Linkurious, provided a web app as a user interface, so that the graph database could be visualised and easily accessed by reporters.

The graph visualisation approach allowed ICIJ journalists to identify the connections between people and bank accounts, helping them 'follow the money' to identify dozens of instances of fraud, corruption and tax evasion.

Below is just one example illustrating these types of graph relationships.



Schema representing the entities and connections within the Swiss Leaks dataset.

## The Results: Unlocking Complex Data for Fraud Investigation

After importing the data into Neo4j and Linkurious, Cabra and her team noted a marked difference in how journalists were able to use the Swiss Leaks data.

"Using Neo4j and Linkurious technology was a complete game-changer for us," Cabra commented. "They enabled our journalists to explore complex financial data in very different ways to before."

The Neo4j-Linkurious solution even met Cabra's other requirement for ease of usability.

"It was so intuitive that very little training was needed to get the reporters up and running," Cabra said. "We just ran a short online demo and webinar. Our reporters were surprised how easy it was to query the data – they didn't need advanced IT skills to do it."

After Cabra's team shared the tool on the ICIJ's virtual newsroom, journalists worldwide tapped into the dataset and the graph analysis tool within their respected regions, querying data on a worldwide scale. By being able to easily visualize the networks around clients and accounts they found many more connections than they had before, which lead to new stories that later made front pages all around the globe. Prior to this, lone reporters had to establish connections by hand with the information of dozens of files – a time-consuming task that could yield inaccurate results.

## CASE STUDY

"It's a revolutionary discovery tool that has transformed our investigative journalism process."

– Mar Cabra

Using Neo4j to investigate the HSBC Leaks data, journalists easily identified major players, intermediaries and beneficiaries in the scandal (regardless of location) and saw how they were connected, in the millisecond click of a mouse.

Since the Linkurious visualization tool was also understandable across cultural and geographic borders, it was the perfect way to achieve international collaboration on the project, proving to be an invaluable tool for reporting as well as for fact-checking as journalists readied their copy.

"It's a revolutionary discovery tool that has transformed our investigative journalism process," Cabra noted. "There is huge potential for us to use graphs to explore other corporate or financial scandals. This simply wouldn't have been possible before on this scale – it's magic!"

Neo4j has now become an integral part of the ICIJ for its Big data projects. Following the Swiss Leaks revelations, some external journalists have also been knocking on the ICIJ's door to use the graph analysis tools for themselves.

### The Aftermath: Unlocking Complex Data for Fraud Investigation

In February 2015, more than 50 news organizations worldwide (including Le Monde) revealed how HSBC had helped criminals, traffickers and tax evaders and profited from doing business with them, by helping shelter over 100,000 clients with accounts worth \$100 billion in Switzerland.

The six-month, 150-journalist project was awarded the prestigious Data Journalism Award (Investigation of the Year category) by the Global Editors Network. The publication of stories is still on-going.

### About The ICIJ: The Fourth Estate's Crack Investigators

The International Consortium of Investigative Journalists (ICIJ) is a global network of more than 190 investigative journalists in more than 65 countries who work together on in-depth investigative stories.

Founded in 1997 by American journalist Chuck Lewis, the ICIJ was launched as a project of the Center for Public Integrity to extend the Center's style of watchdog journalism, focusing on issues like cross-border crime, corruption and the accountability of power. Backed by the Center and its computer-assisted reporting specialists, public records experts, fact-checkers and lawyers, ICIJ reporters and editors provide real-time resources and the latest tools and techniques to journalists around the world.

Neo4j is the leader in graph database technology. As the world's most widely deployed graph database, we help global brands – including [Comcast](#), [NASA](#), [UBS](#), and [Volvo Cars](#) – to reveal and predict how people, processes and systems are interrelated.

Using this relationships-first approach, applications built with Neo4j tackle connected data challenges such as [analytics and artificial intelligence](#), [fraud detection](#), [real-time recommendations](#), and [knowledge graphs](#). Find out more at [neo4j.com](#).

**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)