

## FALLSTUDIE



### Das ICIJ

# Wie das ICIJ Graphdatenbanken für die Swiss-Leaks-Veröffentlichung nutzte

#### BRANCHE

Medien

#### HERAUSFORDERUNG

Unterstützung der Reporter beim Auswerten der komplexen Swiss-Leaks-Daten für besseren investigativen Journalismus

#### STRATEGIE

- Einrichten eines neuen Graphdatenanalyse-Tools, mit dessen Hilfe Journalisten schnell und einfach neue Fakten für ihre Storys finden
- Weltweite, kollaborative Nutzung des Tools beim Swiss-Leaks-Projekt

#### LÖSUNG

- Neo4j war das perfekte Tool für das Verwalten und Abfragen komplexer Daten
- Das Visualisierungs-Tool von Linkurious machte die Datenanalyse für alle ICIJ-Reporter zugänglich

#### ERGEBNIS

- Schnelles und einfaches Durchsuchen und Auswerten komplexer, unstrukturierter Daten für Journalisten weltweit
- Bahnbrechende neue Erkenntnisse und Storys, die sich nur mithilfe der Graphdatenbank aufdecken ließen
- Bahnbrechende, neue Methode zur Betrugsanalyse, durch die der investigative Journalismus insgesamt verändert wird

### Die Macht der Graphen beim investigativen Journalismus

Anfang 2014 gewannen zwei französische Journalisten von Le Monde, Gérard Davet und Fabrice Lhomme, Zugriff auf einen extrem komplexen und zugleich wertvollen Datensatz. Die Daten enthielten Information zu mehr als 100.000 Kunden der Schweizer HSBC-Bank und deren Konten und enthüllten diverse unlautere Machenschaften.

Davet und Lhomme erkannten das Potenzial der Daten für eine bahnbrechende Story über Betrug, Steuerdelikte und internationale Straftaten. Das Problem: Die Daten waren schlicht zu komplex für eine Analyse mit konventionellen Methoden.

### Die Herausforderung: Wertvolle, neue Erkenntnisse aus komplexen Daten ziehen

Üblicherweise müssen Reporter versuchen, Datenzusammenhänge aus Excel-Tabellen heraus zu ziehen, manuelle Internetrecherchen durch zu führen und gelegentlich Verbindungen zwischen Personen und Einheiten heraus zu finden, um die richtigen Fakten für ihre Storys zu erhalten.

Davet und Lhomme stellten jedoch fest, dass der Swiss-Leaks-Datensatz zu umfangreich für eine manuelle, eigenständige Analyse war. Deshalb wendeten Sie sich an das International Consortium of Investigative Journalists (ICIJ), das eine der weitreichendsten bisher dagewesenen journalistischen Zusammenarbeiten startete.

Mar Cabra, Redakteurin der Daten- und Rechercheinheit des ICIJ, war klar, dass für die Analyse der Zusammenhänge dieser Daten für diese und künftige Untersuchungen ein besseres Tool erforderlich war.

Die Swiss-Leaks-Daten beinhalteten Informationen über HSBC-Kontoinhaber in mehr als 200 Ländern, die über eine Summe von insgesamt über 100 Milliarden US-Dollar verfügten. Diese Informationen waren jedoch auf mehrere Tausend Dateien verteilt, die nicht direkt miteinander verbunden waren. Aufgrund der Komplexität der Daten benötigten Cabra und das ICIJ Analysemethoden für die großen Mengen unstrukturierter Daten, um schnell und einfach einen Überblick zu erhalten.

Außerdem war Cabra wichtig, dass keine Unterstützung durch Datenwissenschaftler oder -entwickler benötigt wurde. Cabra wollte, dass der Datenfindungs- und -analysevorgang Reportern weltweit zugänglich war – unabhängig von deren technischem Know-how.

### Die Lösung: Graphanalyse verfolgt Gelder zurück bis zur Quelle

Die Swiss-Leaks-Story war nicht das erste Mal, dass die Organisation mit einem komplexen Datensatz zu tun hatte. Deshalb war Cabra klar, dass das ICIJ auf eine Graphdatenbank-Lösung angewiesen war.

„Bei Storys wie den Offshore-Leaks stellte ich fest, dass die Graphanalyse äußerst wichtig für die Untersuchung von Finanzkriminalität ist“, so Cabra. „Verbindungen sind unerlässlich,

## FALLSTUDIE



„Die Neo4j- und Linkurious-Technologie ermöglichte uns völlig neue Ansätze.“

– Mar Cabra

um hinter den Kern der Story zu kommen: Sie zeigen, wer mit wem Geschäfte tätigt. Bei den HSBC-Leaks entschieden wir uns frühzeitig für einen graphbasierten Ansatz.“

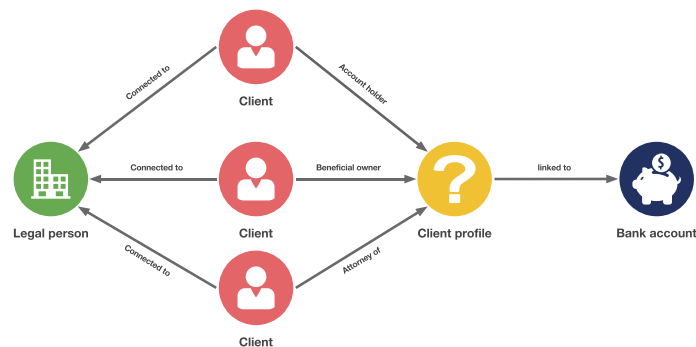
Die Daten- und Rechercheinheit bildete zunächst mithilfe der zur Verfügung stehenden einfachen Excel-Dateien die Datenbank der HSBC-Kunden nach. Anschließend wurde jeder Name mit einem oder mehreren Ländern verbunden (beide werden in der Graphdatenbank als „Knoten“ bezeichnet). Schließlich wurden die Daten in ein Graph-Format umgewandelt, um die Verbindungen zwischen den Knoten zu untersuchen.

Insgesamt lagen ca. 60.000 Dateien vor, die Informationen über mehr als 100.000 Kunden in 203 Ländern beinhalteten. Das Ergebnis war eine Graphdatenbank mit mehr als 275.000 Knoten und 400.000 Verbindungen.

Das ICJ arbeitete mit Open-Source-Integrationssoftware-Anbieter Talend zusammen, um den Originaldatensatz in die Graphdatenbank Neo4j von Neo Technology zu übertragen. Ein weiterer Neo-Partner, Linkurious, stellte eine Web-App als Benutzerschnittstelle bereit, der den Reportern eine Visualisierung und einen einfachen Zugriff auf die Graphdatenbank ermöglichte.

Dank der Visualisierung des Graphen erkannten die ICJ-Journalisten die Verbindungen zwischen Personen und Bankkonten. Dadurch konnten Dutzende Fälle von Betrug, Korruption und Steuerdelikten aufgefunden werden.

Das folgende Beispiel illustriert diese Graph-Zusammenhänge.



Schema mit den Einheiten und Verbindungen im Swiss-Leaks-Datensatz.

### Die Ergebnisse: Durchsuchen komplexer Daten zur Aufdeckung von Betrugsfällen

Nach dem Import der Daten in Neo4j und Linkurious stellten Cabra und ihr Team einen großen Unterschied in den Nutzungsmöglichkeiten der Swiss-Leaks-Daten fest.

„Die Neo4j- und Linkurious-Technologie ermöglichte uns völlig neue Ansätze“, so Cabra. „Unsere Journalisten hatten ganz neue Möglichkeiten bei der Untersuchung komplexer Finanzdaten.“

Die Neo4j-Linkurious-Lösung erfüllte auch Cabras Wunsch nach einer unkomplizierten Anwendung.

„Die Lösung war so intuitiv, dass die Reporter bereits nach einer kurzen Einführung loslegen konnten“, sagt Cabra. „Eine kurze Online-Demo und ein Webinar genügten. Unsere Reporter waren überrascht, wie einfach die Daten abgerufen werden konnten – es waren keine große IT-Erfahrung erforderlich.“

Als Cabras Team das Tool in der virtuellen Nachrichtenredaktion des ICJ veröffentlichte, wurde der Datensatz und das Graphanalyse-Tool von Journalisten in ihren jeweiligen Ländern für die Abfrage weltweiter Daten verwendet. So konnten sie die Netzwerke von Kunden und Konten ganz einfach visualisieren und fanden dadurch weit mehr Verbindungen als zuvor. Dies führte zu neuen Storys, die später weltweit für Schlagzeilen sorgten.

## FALLSTUDIE

“Das Tool hat unsere Abläufe im investigativen Journalismus revolutioniert.”

– Mar Cabra

Früher mussten einzelne Reporter mithilfe Dutzender Dateien die Verbindungen manuell herstellen – eine sehr zeitaufwändige Arbeit, die zu ungenauen Ergebnissen führen konnte. Als die Journalisten für die Untersuchung der HSBC-Leaks-Daten auf Neo4j umstellten, konnten sie in der Millisekunde eines Mausklicks die wichtigsten Akteure, Vermittler und Begünstigten unabhängig von deren Ort ausfindig machen und die Verbindung zwischen ihnen erkennen.

Da das Visualisierungs-Tool Linkurious darüber hinaus über kulturelle und geografische Grenzen hinweg verständlich war, eignete es sich perfekt für die internationale Zusammenarbeit an dem Projekt. Es erwies sich als unentbehrliches Tool für die Berichterstattung und die Faktenüberprüfung bei der Vorbereitung von Manuskripten.

„Das Tool hat unsere Abläufe im investigativen Journalismus revolutioniert“, betont Cabra. „Für die Verwendung von Graphen besteht auch bei anderen Unternehmens- und Finanzskandalen ein großes Potenzial. Früher wäre das in diesem Umfang nicht möglich gewesen – es ist unglaublich.“

Inzwischen ist Neo4j Teil aller ICIJ Projekte, die große Datenmengen mit sich bringen. Seit den Swiss-Leaks-Enthüllungen wird das ICIJ immer wieder von externen Journalisten kontaktiert, die das Graphanalyse-Tool ebenfalls verwenden möchten.

### Die Folgen: Durchsuchen komplexer Daten zur Untersuchung von Betrugsfällen

Im Februar 2015 enthüllten mehr als 50 Nachrichtenagenturen weltweit (darunter Le Monde), wie HSBC Straftaten wie Drogen- und Menschenhandel und Steuerdelikte unterstützte und davon profitierte, indem sie mehr als 100.000 Kunden und deren Konten im Wert von 100 Milliarden US-Dollar in der Schweiz verschleierte.

Das sechsmontatige Projekt, an dem 150 Journalisten beteiligt waren, erhielt den prestigeträchtigen Data Journalism Award in der Kategorie „Investigation of the Year“ (Recherche des Jahres) des Global Editors Network. Die Veröffentlichung von Storys aus der Recherche ist noch nicht abgeschlossen.

### Zum ICIJ: Die Investigativ-Experten der Medienbranche

Das ICIJ (International Consortium of Investigative Journalists) ist ein globales Netzwerk von mehr als 190 investigativen Journalisten in mehr als 65 Ländern, die gemeinsame Recherchen zu Storys betreiben.

Die Organisation wurde 1997 von US-Journalist Chuck Lewis zunächst als Projekt des Center for Public Integrity gegründet, um den unabhängig ermittelnden Journalismus auszuweiten. Der Fokus lag dabei auf grenzüberschreitenden Straftaten, Korruption und Machtmissbrauch. Die ICIJ-Reporter und -Redakteure stellen Journalisten weltweit Echtzeit-Ressourcen und die neuesten Tools und Methoden zur Verfügung, in Zusammenarbeit mit dem Center und dessen rechnerunterstützte Reporter, Experten für öffentliche Datensätze und Faktenuntersuchung sowie Juristen.

[www.icij.org](http://www.icij.org)

#### Info zu Neo Technology

Neo Technology entwickelt die weltweit führende Graphdatenbank Neo4j, mit der sich Datenverbindungen optimal nutzen lassen. Dazu zählen Anwendungen wie das Angebot von personalisierten Empfehlungen für Produkte und Dienstleistungen, zusätzliche soziale Medienfunktionen für Websites, Analyse von Telekommunikationsnetzwerken oder die Neuorganisation von Stammdatenverwaltung sowie Identity und Access Management. Unternehmen nutzen Graphdatenbanken, um Datenvernetzungen und -zusammenhänge abbilden, abspeichern und abfragen zu können. Großunternehmen wie Walmart, eBay, UBS, Nomura, Cisco, HP und Telenor sowie Startups wie CrunchBase, Medium, Polyvore und Zephyr Health verwenden Neo4j, um erfolgsentscheidenden Nutzen aus Datenverbindungen zu ziehen.

**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)