# « Panama Papers » : le big data au service du journalisme d'investigation

## Description

**Si les « Panama Papers » ont révélé l'ampleur de la fraude fiscale et du blanchiment d'argent opérés dans une cinquantaine de pays, ils ont également mis en lumière la nécessité pour les journalistes de faire collaborer leurs rédactions et de s'emparer d'outils technologiques capables de traiter et d'analyser un très grand nombre de données.**

Les « Panama Papers » sont la plus grosse fuite d'informations jamais exploitée par des médias. Tout au long de l'année 2015, un lanceur d'alerte, sous le pseudonyme de John Doe, a envoyé au quotidien allemand *Süddeutsche Zeitung* près de 11,5 millions de documents confidentiels provenant des archives du cabinet d'avocat panaméen Mossack Fonseca, accumulées entre 1977 et 2015.

Ces documents détaillent le montage financier de quelque 214 000 sociétés *offshore* ainsi que de leurs 450 000 actionnaires, parmi lesquels 12 chefs d'Etat et 143 hommes politiques de plus de 50 pays, des célébrités ou encore des sportifs de haut niveau. La totalité de ces 11,5 millions de documents représente 2,6 téraoctets (2 600 Go) de données soit plus que les fuites réunies de Wikileaks (2010), Offshore Leaks (2013), LuxLeaks (2014) et Swiss Leaks (2015). Ce sont 4,8 millions de courriels, 3 millions de bases de données, 2 millions de fichiers PDF, 1,1 million d'images (dont les photocopies des passeports des actionnaires et les scans de contrats signés), 320 000 fichiers « texte » et environ 2 000 fichiers d'autres formats.

Si les sociétés *offshore* ne sont pas illégales dans la plupart des pays, cette fuite révèle en revanche l'ampleur de l'évasion fiscale et du blanchiment d'argent, mis en œuvre à travers des sociétés-écrans pour le compte de clients, par les plus grands établissements bancaires comme la Deutsche Bank, HSBC, la Société générale ou encore Commerzbank ; ce qui, pour le coup, est totalement illégal.

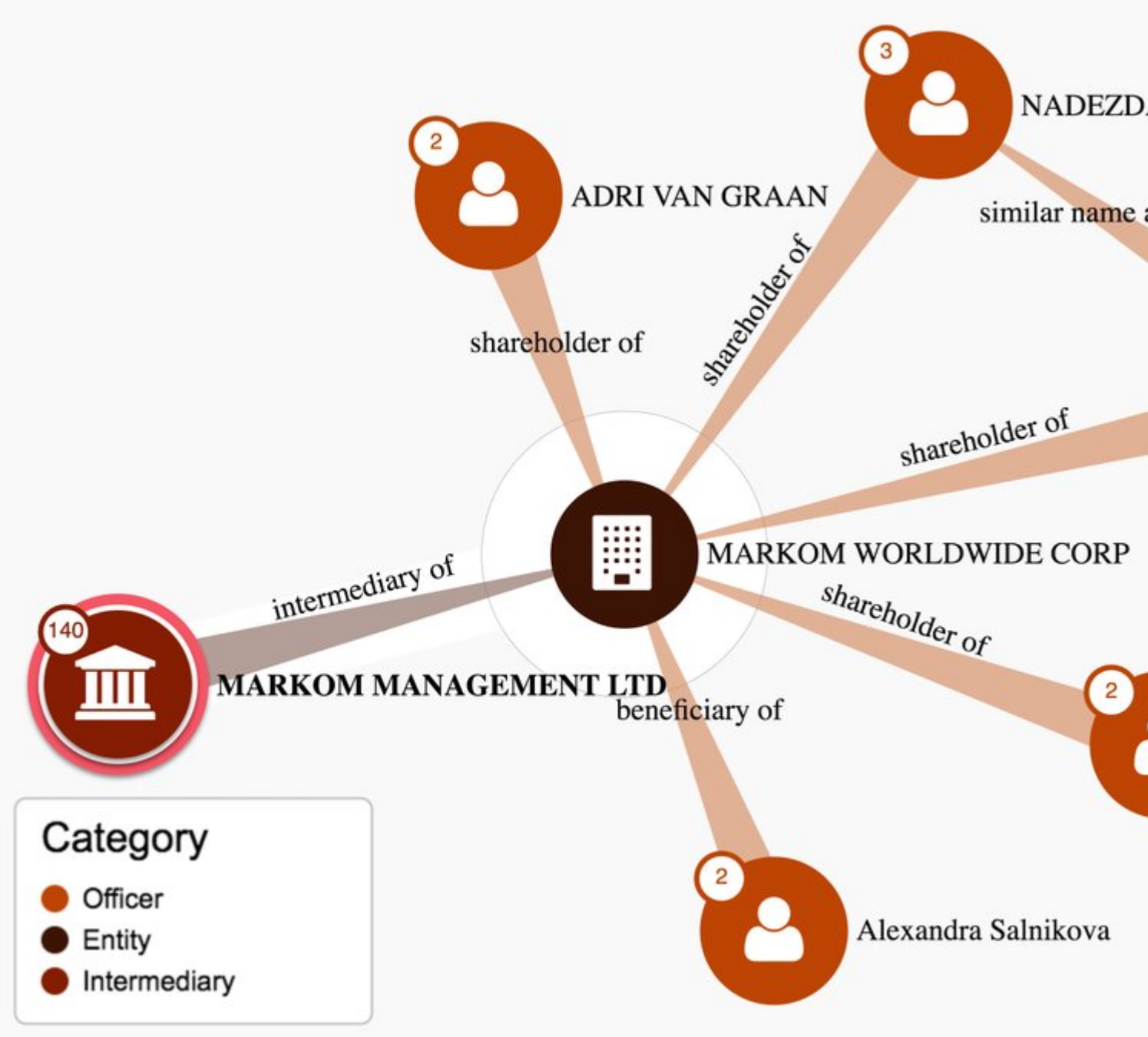
Au-delà de cette avalanche impressionnante de chiffres, le traitement d'une telle masse d'informations a nécessité du quotidien allemand qu'il s'associe au Consortium international des journalistes d'investigation (CIJI) basé à Washington, déjà impliqué lors de l'Offshore Leaks, du LuxLeaks ou encore du Swiss Leaks. Par leur intermédiaire, plus de 107 médias répartis dans 76 pays se sont associés pour analyser et décrypter ces documents pendant un an. Mais les 370 journalistes n'auraient jamais pu les parcourir un à un.

Le Consortium s'est donc appuyé sur un certain nombre d'outils technologiques afin de répondre à deux principaux enjeux. Le premier était celui de la sécurité. Mar Cabra, Data and Research Unit Editor au sein de l'ICIJ explique ainsi que « *nous voulions protéger l'identité des journalistes et nous assurer que nos données ne tombent pas dans les mauvaises mains. Ce qui était particulièrement difficile dans la mesure où nous allions exposer des personnalités politiques dans des pays où la liberté de la presse n'est pas toujours garantie*

». Le second enjeu était celui de pouvoir traiter et analyser cet « océan de données ».

Le Consortium a d'abord utilisé des logiciels de reconnaissance optique de caractères, permettant de convertir en fichier texte des images ou des scans, afin de pouvoir les intégrer dans une base de données interrogeable. Ces données ont ensuite été indexées, notamment grâce au logiciel Nuix, développé par une entreprise australienne dont l'objet est « *d'indexer, explorer, analyser et extraire la connaissance à partir de données non structurées* ». Le Consortium a également mis à la disposition des journalistes travaillant sur l'enquête un moteur de recherche couplé à une plate-forme de visualisation de données, fourni par la start-up française Linkurious.

Fondée en 2013 par Jean Villedieu et Sébastien Heymann, qui s'était déjà distingué comme étant le coauteur de Gephi, une référence dans les technologies d'analyse de graphes *open source*, l'entreprise propose de « *mettre l'analyse de graphes entre les mains des analystes métiers* ». Un graphe est un ensemble de points nommés nœuds reliés par des traits (segments) ou flèches nommées arêtes (ou liens ou arcs), l'ensemble des arêtes entre nœuds forme une figure similaire à un réseau. Ainsi, plutôt que de rechercher des données en tapant une requête qui affichera une liste de résultats, l'outil de Linkurious permet « *à partir d'un mot clé, de remonter au(x) nœud(s) correspondant(s)* ».




Source : https://offshoreleaks.icij.org/nodes/10119701

Ainsi, l'exploration des données est bien plus efficace qu'en s'appuyant sur une base de données

traditionnelle. Depuis 2013, Linkurious a fourni d'importants efforts pour offrir une ergonomie permettant à des non-spécialistes d'utiliser un tel outil, et c'était bien là le souhait du Consortium de fournir un logiciel « *afin que tous les journalistes impliqués de par le monde puissent travailler sur ces données, peu importe leur niveau de maîtrise technique* ».

Le Consortium a donc commencé par extraire les métadonnées des documents avant de relier ces données entre elles pour modéliser dans un graphe un réseau de points et de relations ; ensuite la technologie de Linkurious a permis aux équipes de journalistes d'investigation de chaque pays impliqué d'explorer les données selon leur souhait. Le *big data* au service du journalisme.

Ainsi, les équipes du journal *Le Monde* ont croisé la base de données du Consortium avec la liste complète des « *parlementaires français et européens, la liste Challenges des 500 Français les plus riches (et son corollaire à l'international, la liste Forbes), les organigrammes des principaux partis politiques français, la liste des ministres français depuis les années 1980, les administrateurs du CAC 40, les personnalités préférées des Français, les personnes citées dans des affaires politico-judiciaires depuis 2000, les chefs d'Etat et de gouvernement du monde entier ou encore les joueurs de l'équipe de France de football* ».

A ce jour, les « Panama Papers » révèlent que 1 000 noms sont liés à une adresse en France. Le Service de traitement des déclarations rectificatives (STDR), cellule de régularisation des exilés fiscaux, est d'ailleurs submergée d'appels, au point de passer d'une équipe de 159 à 209 agents à partir de juin 2016.

Dans un souci de transparence, le Consortium a décidé de publier une partie des « Panama Papers ». Le 9 mai 2016, des données structurées recensant le nom de 200 000 sociétés offshore et de leurs actionnaires ont déjà été rendues accessibles sur le site Offshoreleaks.icij.org. En effet, « *il y a tellement de documents. C'est impossible de savoir ce que vous ratez jusqu'à ce que vous trouviez quelque chose* », confesse Gerard Ryle, directeur du Consortium international des journalistes d'investigation (CIJI).

Sources :


- « Panama papers : comment *Le Monde* a travaillé sur plus de 11 millions de fichiers », Maxime Vaudano et Jérémie Baruch, Lemonde.fr, 3 avril 2016.
- « Panama Papers. La plus grande fuite de l'histoire du journalisme », Lucie Geffroy, Courrierinternational.com, 4 avril 2016.
- « Linkurious, la start-up du Big Data qui surfe sur les Panama Papers », Reynald Fléchaux, Silicon.fr, 6 avril 2016.
- « « Panama papers » : un défi technique pour le journalisme de données », Jérémie Baruch et Maxime Vaudano, datajournalistes aux Décodeurs, Data.blog.lemonde.fr/, 8 avril 2016.
- « Y aura-t-il un effet « Panama papers » sur les régularisations auprès de Bercy ? », Mathilde Damgé, Lemonde.fr/, 14 avril 2016.
- Panama Papers, Wikipedia, Fr.wikipedia.org/wiki/Panama_Papers, consulté le 18 mai 2016.

- Offshore Leaks Database, https://offshoreleaks.icij.org, consulté le 18 mai 2016 The International Consortium of Investigative Journalists, Panamapapers.icij.org, consulté le 18 mai 2016.
- Linkurio.us http://linkurio.us/blog, consulté le 18 mai 2016.

**Categorie**

1. Infographies
2. Techniques

**date créée**
19 juillet 2016
**Auteur**
jacquesandrefines