

Linkurious, la start-up du Big Data qui surfe sur les Panama Papers

Les « Panama Papers », cette série de révélations sur la fraude fiscale issues de l'exploitation d'une gigantesque fuite de données au sein du cabinet d'avocats panaméen Mossak Fonseca, ont déjà fait leurs premières victimes, comme le Premier ministre islandais. Mais la société française Linkurious (dont l'équipe est en photo ci-dessus) a, elle, tout à gagner de la proportion que prennent ces révélations. Car la technologie de cette start-up, fondée en 2013 par **Sébastien Heymann**, qui avait participé à la création de [Gephi](#), un moteur Open Source de visualisation de données téléchargé plus d'un million de fois, et par **Jean Villedieu**, a été employée par une partie des 370 journalistes qui ont été mobilisés, dans le monde entier, pour exploiter **11,5 millions de documents**. Une somme d'informations qu'un lanceur d'alertes a transmis au quotidien allemand *Süddeutsche Zeitung*, qui, devant l'ampleur de la tâche, s'est associé à une centaine de médias internationaux mobilisés par le consortium international des journalistes d'investigation, l'ICIJ.

Ce sont ces 2,6 To de données, portant sur 214 488 structures offshore, que la solution de dataviz de Linkurious a aidé à explorer. *« Le premier défi pour l'ICIJ et le Süddeutsche Zeitung a consisté à rendre cette masse de données exploitable pour leur réseau de journalistes d'investigation dans le monde entier, écrit Sébastien Heymann, l'un des co-fondateurs de Linkurious, dans [un billet de blog](#). Ce qui implique de transformer 11,5 millions de documents non structurés en quelque chose d'intelligible pour les journalistes ».*

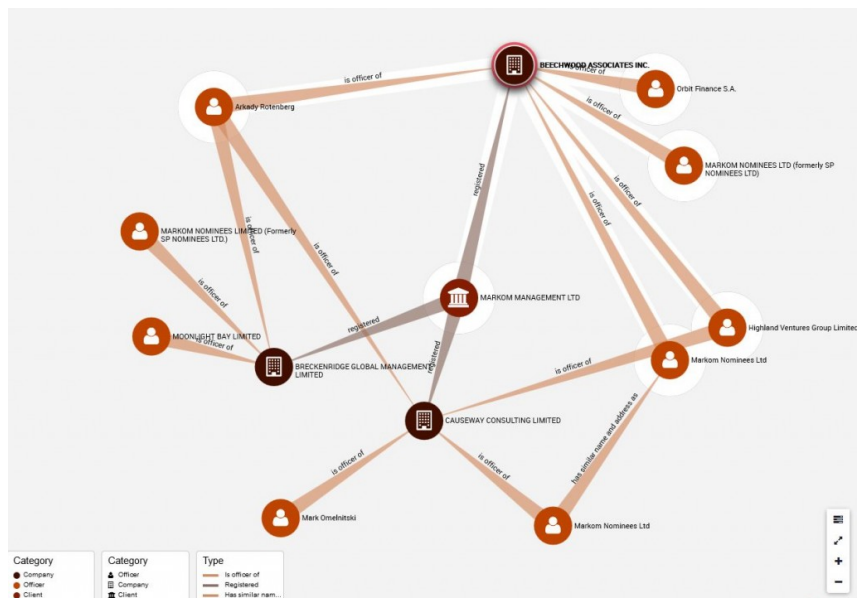
Linkurious pour collaborer

Les spécialistes du département Data & Research de l'ICIJ ont commencé par extraire les métadonnées des documents récupérés, via les outils Apache Solr et Tika, puis ont relié les données entre elles pour créer un réseau de points et de relations, modélisé dans un graphe. Les données ont été **stockées dans Neo4j**, une base de données éditée par Neo Technology. *« Une fois la base créée, ICIJ a pu commencer à la rendre accessible aux équipes d'investigation dans le monde entier »,* détaille Sébastien Heymann. Au travers d'un outil d'exploration permettant de mettre en évidence les réseaux de relations que renferment les documents : Linkurious Enterprise.

Selon le co-fondateur de la start-up, cet outil d'exploration de données a également été exploité par les 370 journalistes travaillant sur l'affaire pour partager des visualisations et collaborer, en toute sécurité. Sensible par nature, l'enquête sur les Panama Papers devait rester secrète jusqu'à sa révélation dans les journaux et **protéger l'identité des journalistes mobilisés**, en particulier de ceux travaillant dans des pays où la liberté de la presse n'est pas garantie.

Retenu pour les Swiss Leaks

Notons que l'ICIJ a, [sur son site](#), largement mis à profit la technologie de dataviz pour présenter les premiers résultats des Panama Papers publiés à travers le monde. Dans son billet de blog, Sébastien Heymann explique d'ailleurs que l'ensemble des Panama Papers – comprendre la partie des données qui sera exploitée par les journaux partenaires – sera présenté sous forme de graphiques Linkurious à partir de début mai.



La technologie de la start-up avait déjà été exploitée par le Consortium international des journalistes d'investigation (ICIJ) lors d'[une précédente enquête](#) (les Swiss Leaks, soit l'exploitation des données dérobées à HSBC par Hervé Falciani). A l'époque déjà, deux journalistes du *Monde*, Gérard Davet et Fabrice Lhomme, s'étaient tournés vers l'ICIJ, en raison du volume de données à traiter et du caractère international de l'affaire, portant sur environ 100 000 clients de HSBC. Le consortium international avait alors choisi Neo4j et Linkurious pour consolider les données et les présenter sous forme de graphes.

A lire aussi :

[Big Data : les entreprises ne sont pas au niveau](#)

[Tour d'horizon des 11 technologies prometteuses du Big Data](#)