

Les Panama Papers

PHP Tour 2017

Bonjour a tous !

Benoit Simard



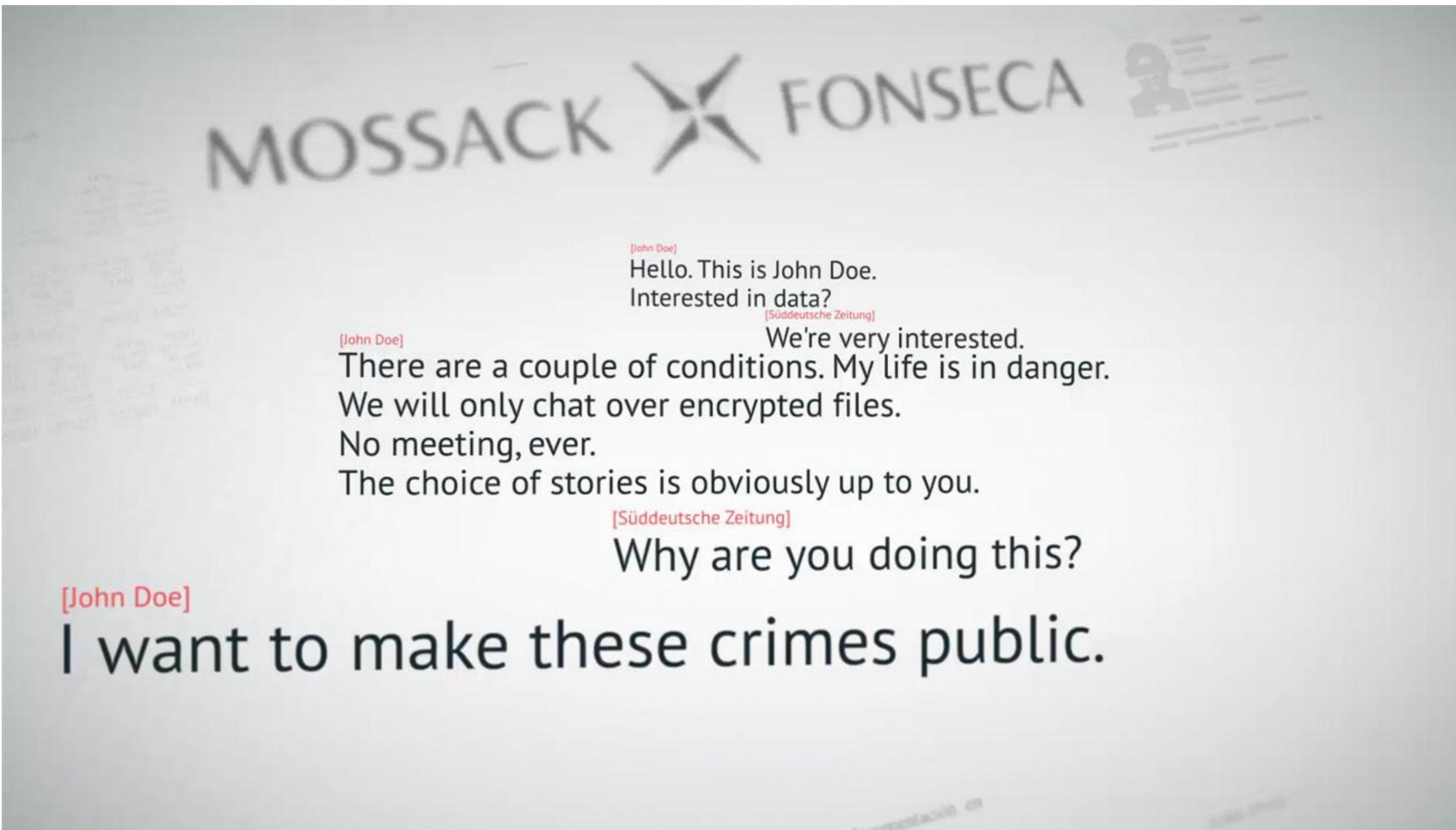
- Consultant Neo4j
- Addicte aux graphes
- Libérateur de données
- @logisima
- benoit@neo4j.com



Revenons 2 ans en arrière.

L'origine

En 2015, Un lanceur d'alerte partage des données avec le journal allemand **Süddeutsche Zeitung**



La volumetrie

C'est la plus grosse fuite de données de l'histoire, et de loin !

11,5 millions de documents représentant **2,6To**





International Consortium of Investigative Journalists



~200 journalistes dans plus de **65 pays**

Equipe de **12 personnes à temps plein** (USA, Costa Rica, Venezuela, Germany, France, Spain)

50% travaillent sur l'analyse des données



Le bilan

Révélation des Panama Papers en Avril 2016, et ouverture des données en mai.

Un impact mondial, et qui continue toujours, avec l'implication de

- **12** chefs d'états
- **128** dirigeants politiques ou hauts fonctionnaires
- **14 000** sociétés, banques ou cabinets d'avocats
- **511** banques liées directement
- **150** procédures judiciaires dans **18** pays

Les fondateurs de **Monsack Fonseca** ont été arrêtés en Février pour blanchiment d'argent.

Prix Pulitzer

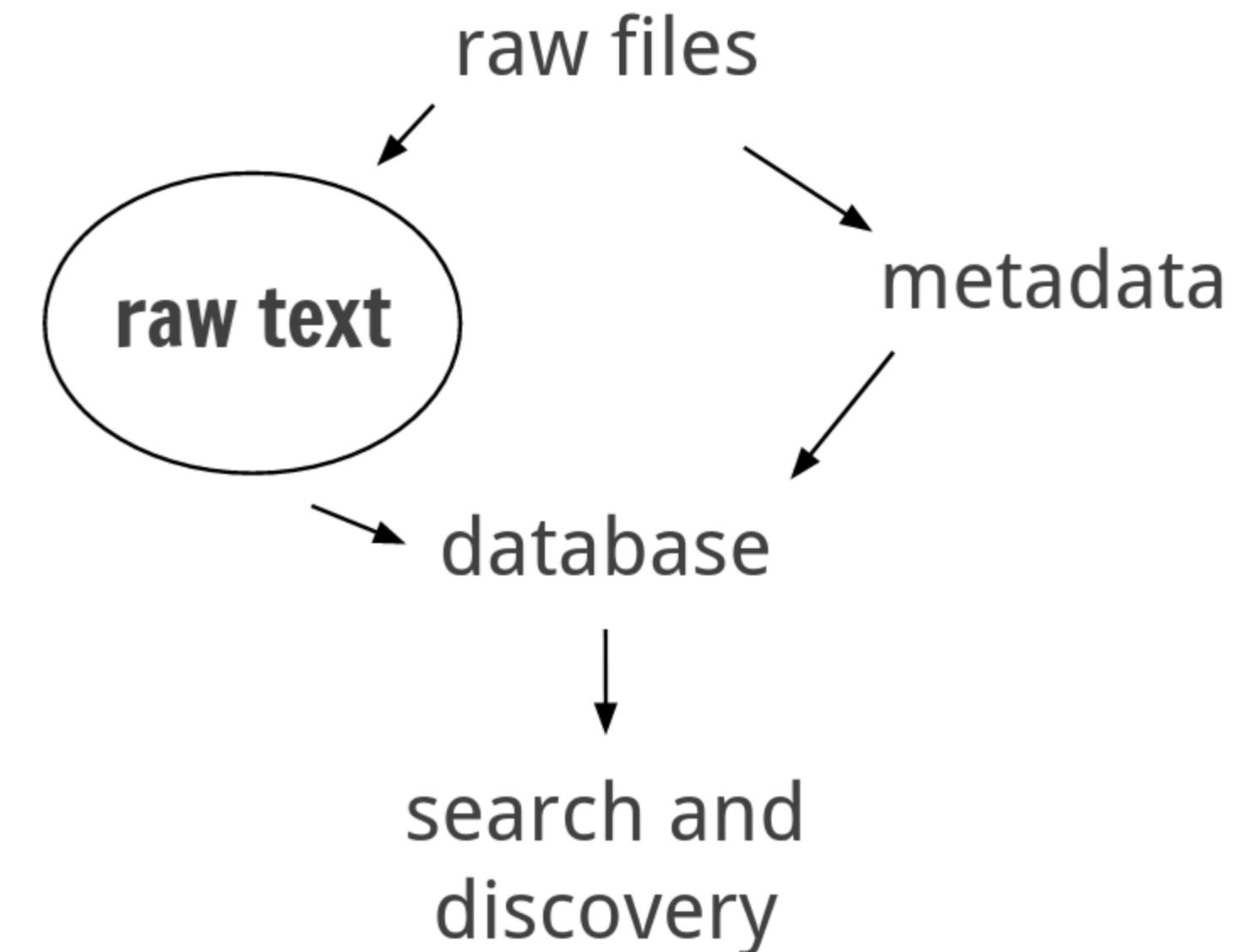
10 Avril 2017

Les panama papers ont gagné le prix Pulitzer pour leur investigation et l'utilisation d'outils novateurs avec la collaboration de plus 300 journalistes repartis sur 6 continents.



Comment ils ont fait ?

Le processus



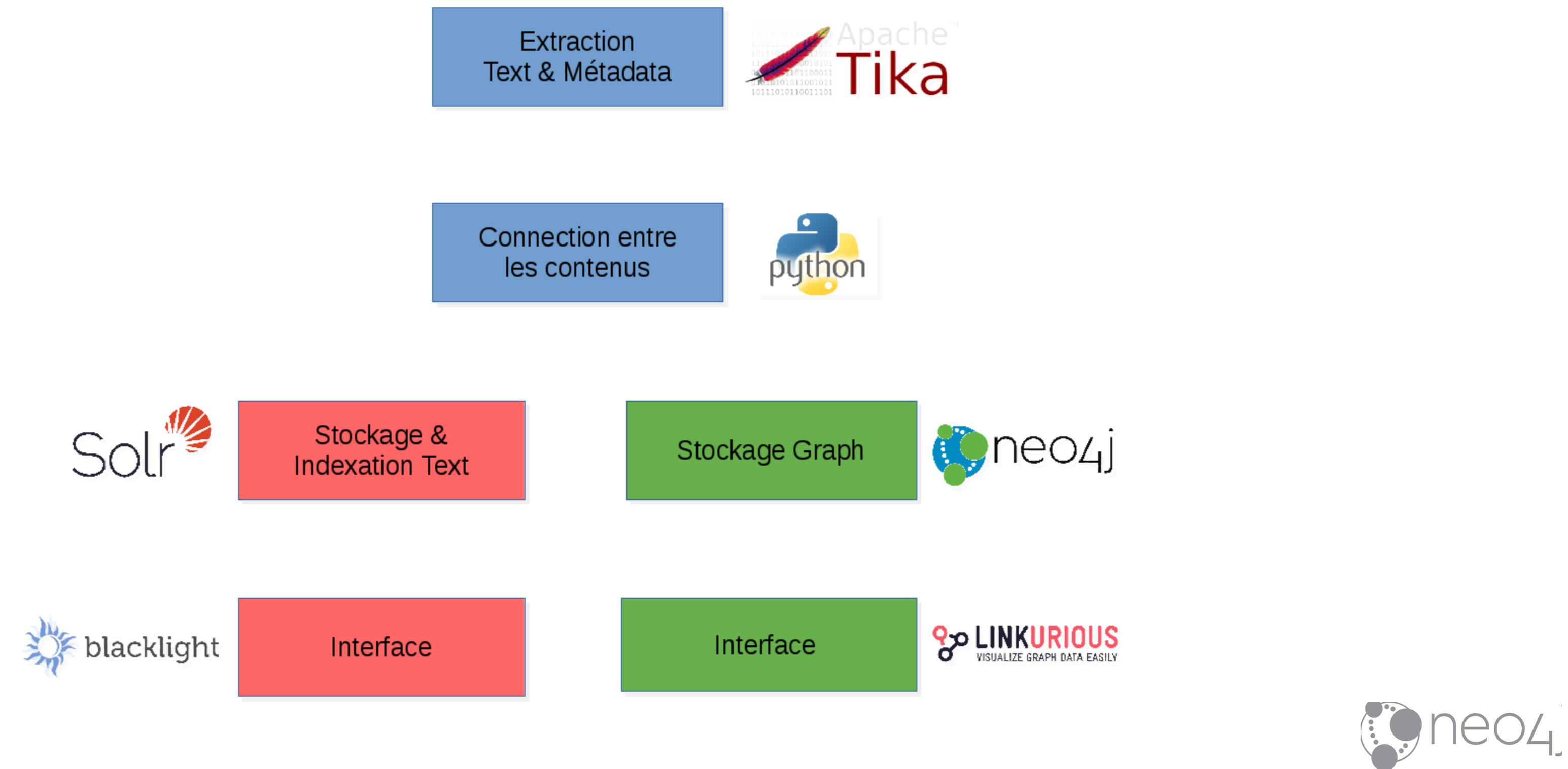
OCR et Cloud

3 millions de fichiers x 10 secondes/fichier = 1 année

1 année / 35 serveurs = 1,5 semaine



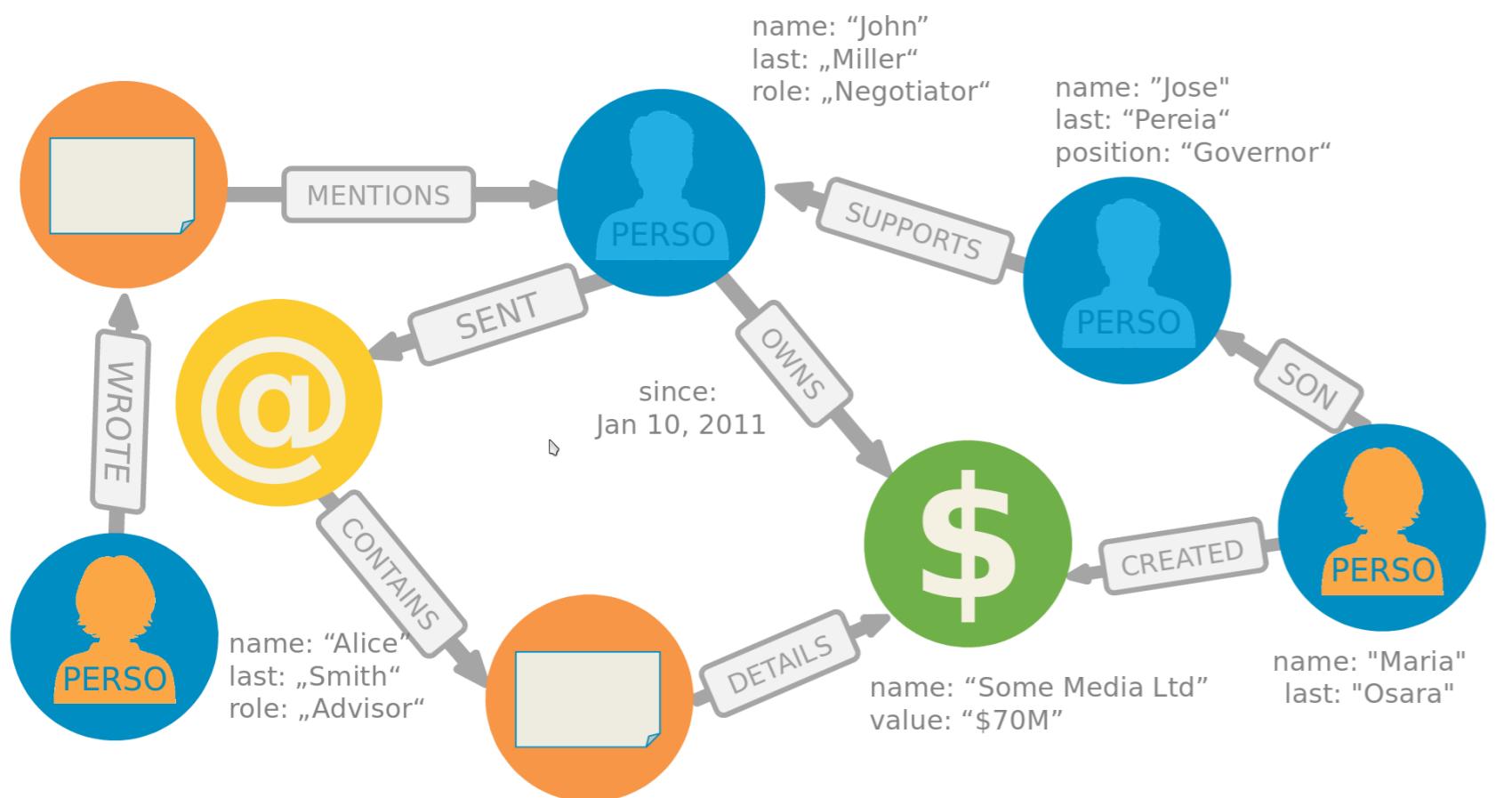
La stack technique



Neo4j



Les graphes de propriétés



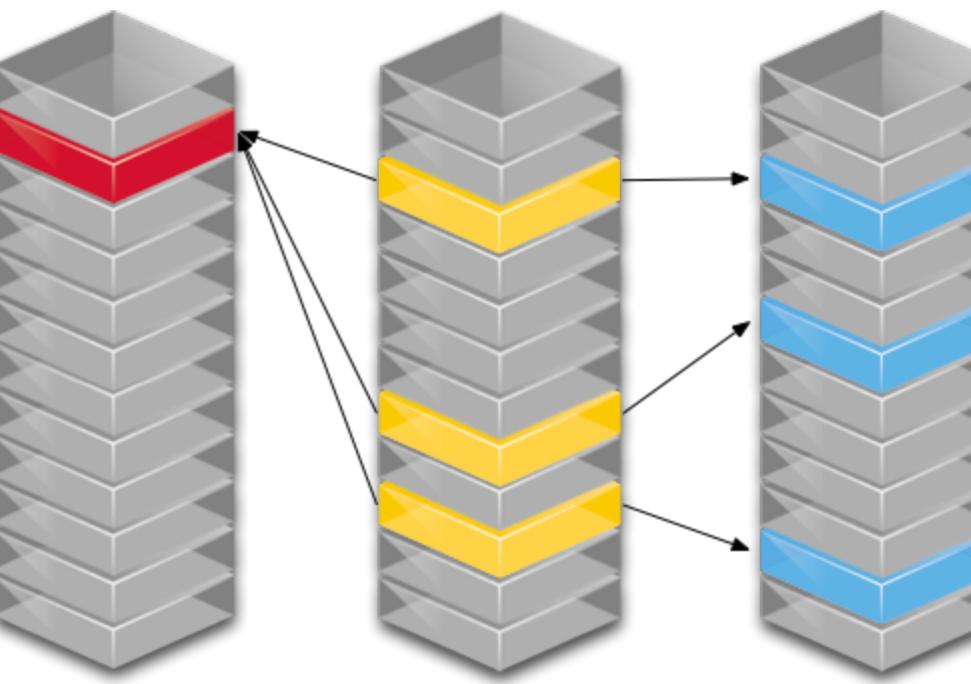
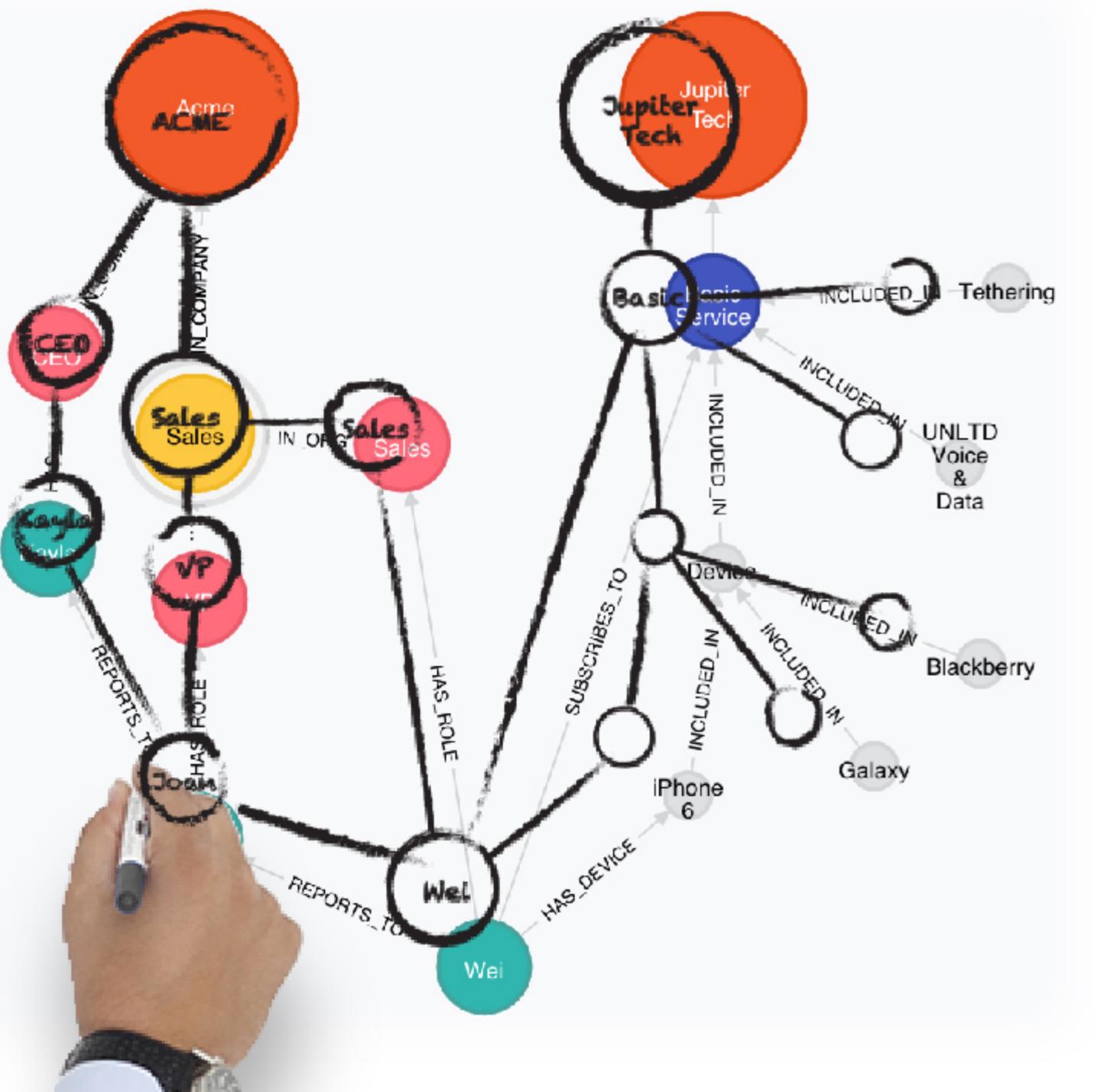
Les noeuds

- Les entités du graphe
- Peuvent avoir des propriétés
- Peuvent avoir des labels (étiquettes)

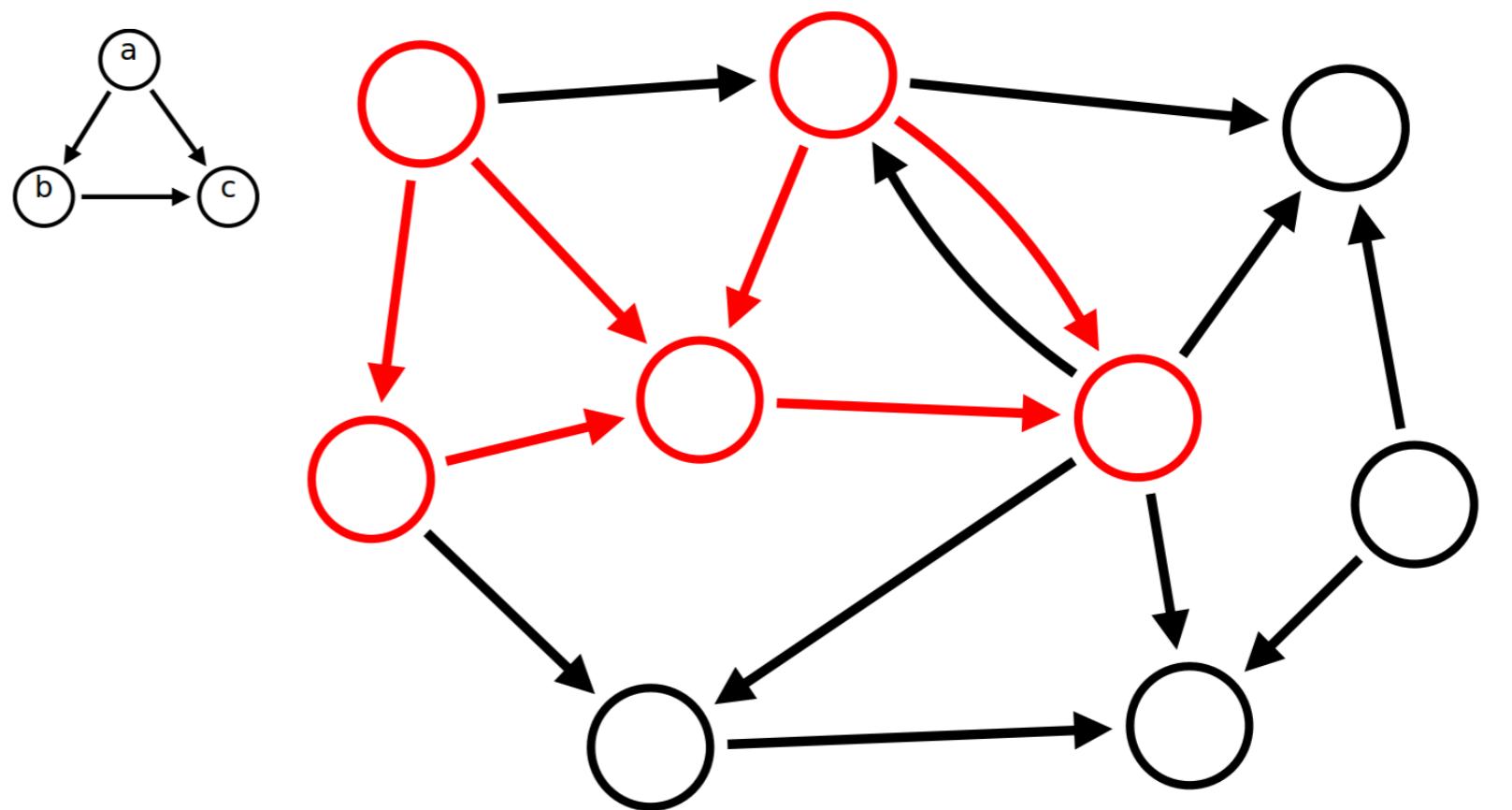
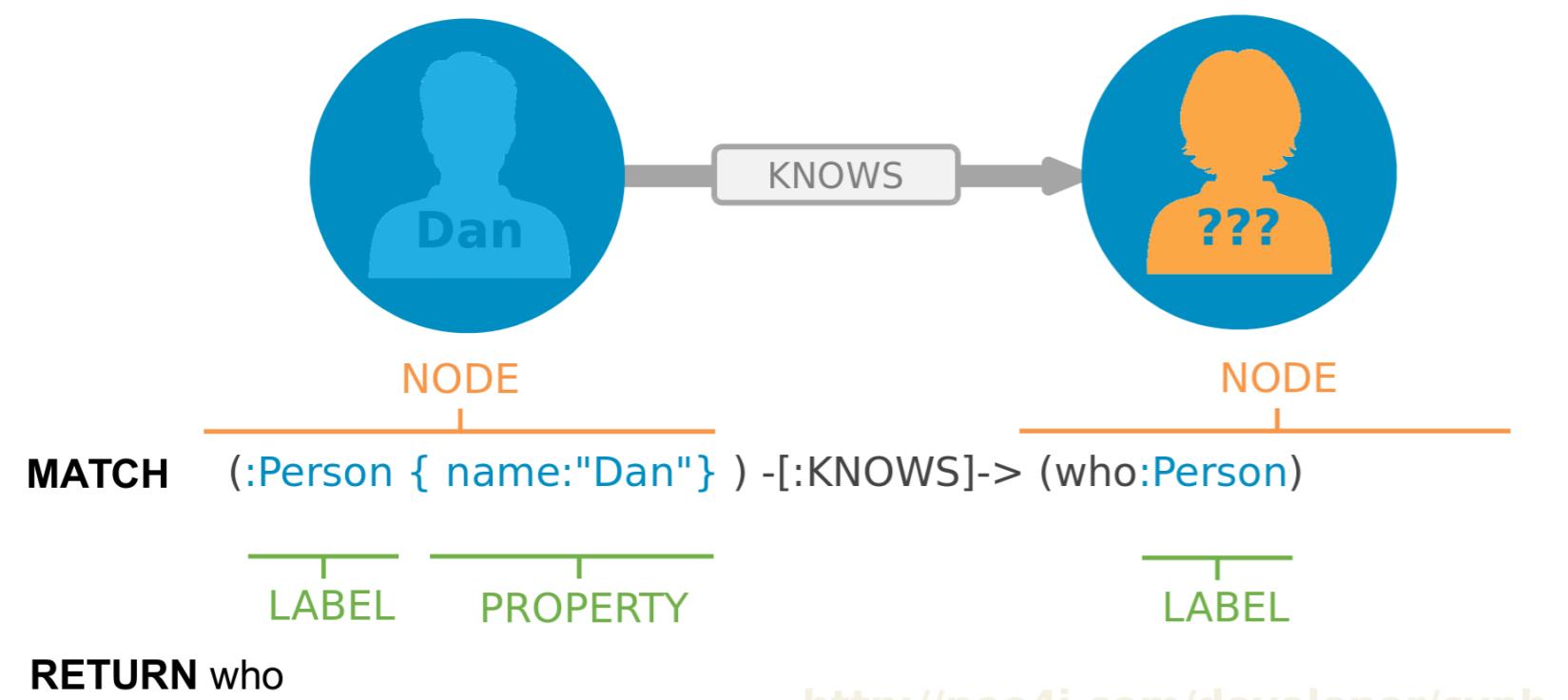
Les relations

- Relient des noeuds avec un type et une direction
- Peuvent avoir des propriétés

Des relations locales

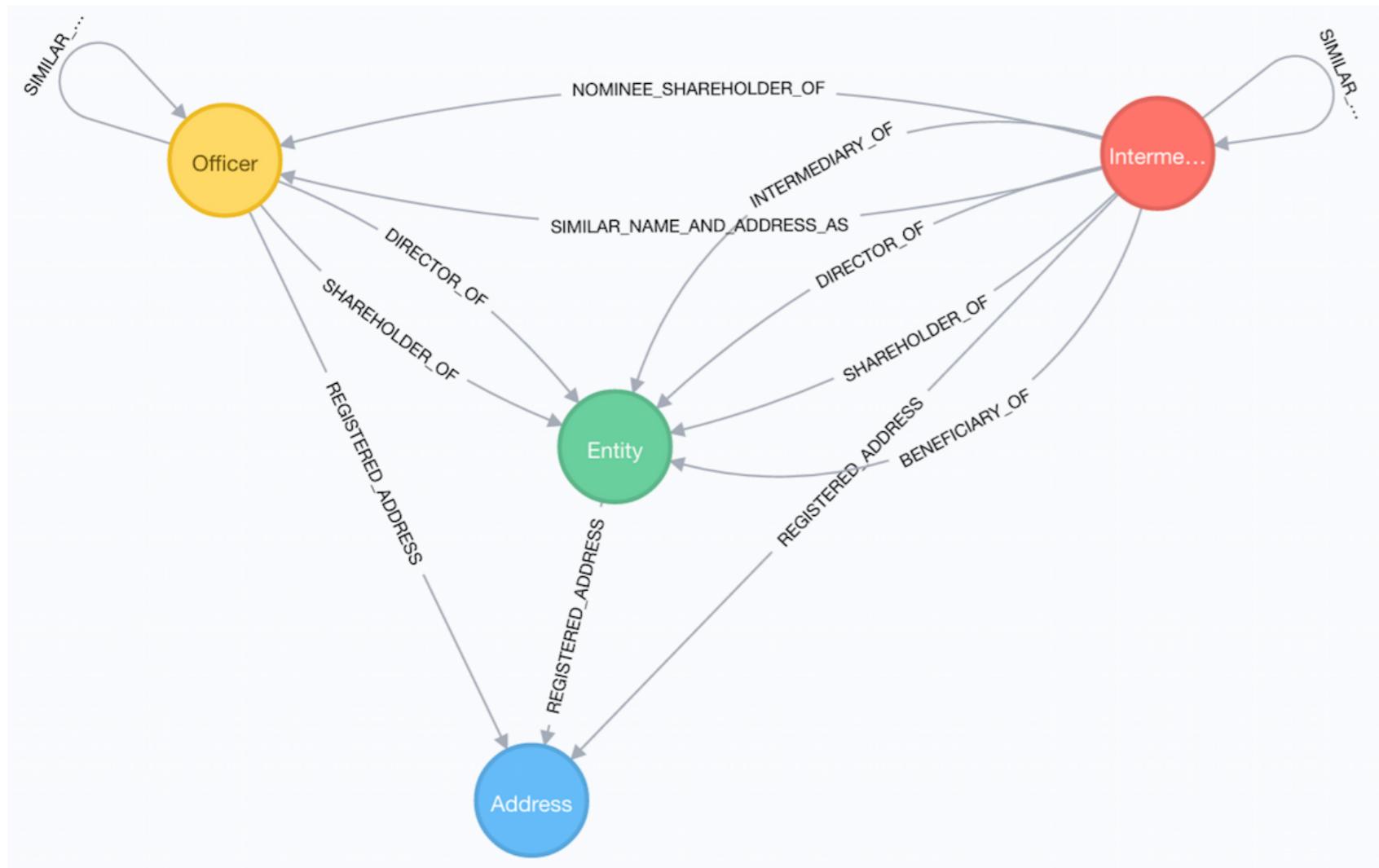


Cypher



Revenons au Panama Papers

Le modèle



- **Entity** : L'entité offshore
- **Officer** : Une personne (morale ou physique) ayant un lien avec l'entité offshore
- **Intermediary** : Un intermédiaire, généralement un cabinet d'avocat
- **Address** : Une adresse

Quelques stats

```
1 MATCH (n:Entity) RETURN count(*);  
2 MATCH (n:Officer) RETURN count(*);  
3 MATCH (n:Intermediary) RETURN count(*);  
4 MATCH (n:Address) RETURN count(*);
```

Le top 20 des intermédiaires

```
1 MATCH (i:Intermediary) WHERE size( (i)--() ) > 100
2 MATCH (i)-[:INTERMEDIARY_OF]->(entity:Entity)
3 RETURN i.name as intermediary, count(*) as count
4 ORDER BY count DESC LIMIT 20
```

Mossack Fonseca

Regardons les entreprises offshores créées par Mossack Fonseca

```
1 MATCH (i:Intermediary)-[r:INTERMEDIARY_OF]->(e:Entity)
2 WHERE i.name CONTAINS "MOSSACK FONSECA"
3 RETURN i, r, e LIMIT 100
```

Les paradis fiscaux populaires des français

```
1 MATCH (o:Officer)-->(e:Entity)<-[:INTERMEDIARY_OF]-(i:Intermediary)
2 WHERE o.country_codes CONTAINS 'FRA'
3 RETURN e.jurisdiction_description AS jurisdiction, count(*) AS number
4 ORDER BY number DESC LIMIT 10
```

Les intermédiaire populaire a Paris

```
1 MATCH (a:Address)<-[:REGISTERED_ADDRESS]-(o:Officer),  
2 (o)-->(e:Entity)<-[ :INTERMEDIARY_OF ]-(i:Intermediary)  
3 WHERE a.address CONTAINS 'Paris' AND a.address CONTAINS 'France'  
4 RETURN i.name AS intermediary, count(DISTINCT e) AS number  
5 ORDER BY number DESC LIMIT 10
```

Les couples gagnants

En investigation, il est important d'identifier les gens qui opèrent ensemble

```
1 MATCH (o1:Officer)-[r1]->(e:Entity)<-[r2]-(o2:Officer)
2 WHERE id(o1) < id(o2)
3 AND size( (o1)-->() ) > 2 AND size( (o2)-->() ) > 2
4 WITH o1, o2, count(*) as freq, collect(e.name)[0..10] as entities
5 WHERE freq > 2
6 RETURN o1.name, o2.name, freq, entities
7 ORDER BY freq DESC
8 LIMIT 10
```

Ces petits chemins ...

Mais aussi de voir les liens possible entre deux personnes

```
1 MATCH (a:Officer),(b:Officer)
2 WHERE a.name CONTAINS 'Smith'
3   AND b.name CONTAINS 'Grant'
4 MATCH p=allShortestPaths((a)-[:OFFICER_OF|:INTERMEDIARY_OF|:REGISTERED_ADDRESS*..10]-(b))
5 RETURN p
6 LIMIT 50
```



Pour aller plus loin

Panama sandbox

<https://neo4j.com/sandbox-v2>

Recommendations



Generate personalized real-time recommendations using a dataset of movie reviews.

[Launch Sandbox](#)

Trumpworld



Explore connections in and around the Trump Administration using this dataset from BuzzFeed.

[Launch Sandbox](#)

Twitter



If signed into Neo4j Sandbox using Twitter, this Sandbox will allow you to Graph your Twitter network.

[Launch Sandbox](#)

Legis-Graph



US Congress modeled as a Graph - bills, votes, members, and more.

[Launch Sandbox](#)

Network and IT Management



Dependency and root cause analysis + more for network and IT management

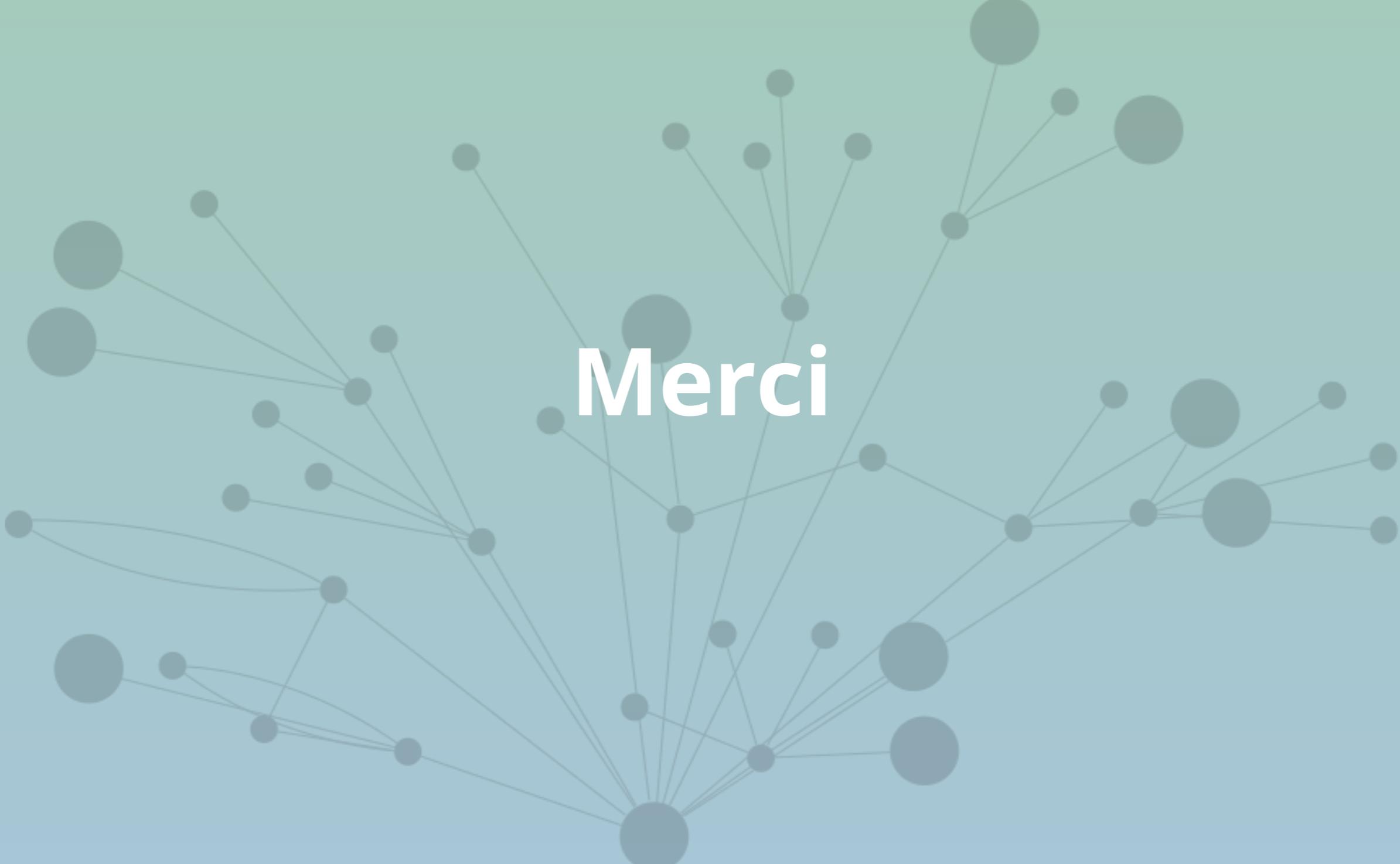
[Launch Sandbox](#)

Panama Papers by ICIJ



The Panama Papers dataset and guide from the International Consortium of Investigative Journalists (ICIJ).

[Launch Sandbox](#)



Merci

Des questions ?



En savoir plus

- <http://neo4j.com/blog/panama-papers/>
- <https://panamapapers.icij.org/>
- <http://panamapapers.sueddeutsche.de/en/>
- <http://www.theguardian.com/news/series/panama-papers>