

Cour des comptes



Chambres régionales
& territoriales des comptes

LA MISE EN ŒUVRE PRATIQUE DES TECHNOLOGIES DE L'INFORMATION DANS LES INSTITUTIONS RÉGIONALES DE CONTRÔLE

Xavier BAILLY, conseiller de la Chambre régionale des comptes Occitanie

Octobre 2017



0.1 – Les technologies de l’information modifient la façon de travailler des juridictions financières françaises

- Les technologies de l’information modifient en profondeur, et de plusieurs manières la façon de travailler des juridictions financières françaises. Le basculement des chambres régionales des comptes vers le numérique autorise ainsi deux usages différenciés :
- L’aide à la programmation des contrôles, correspondant à l’utilisation du système par la présidence et l’encadrement des chambres ;
- L’appui au contrôle, correspondant à une utilisation pour les instructions menées par les équipes de contrôle des chambres ;
- Il permet également une approche différente de la communication une fois les rapports publiés.



0.2 – L'émergence d'une stratégie fondée sur l'analyse des données

- Une pratique s'appuyant sur quelques outils plus anciens (2007-2012) de suivi de l'activité ;
- Une **stratégie récente** (2012-2014) avec de nombreux défis :
 - La collecte des données dont les juridictions ne sont pas propriétaires ;
 - Leur extraction et leur transformation ;
 - La manipulation des données, leur analyse,
 - La visualisation,
 - La transformation des données analysées en « histoire » publiable ;

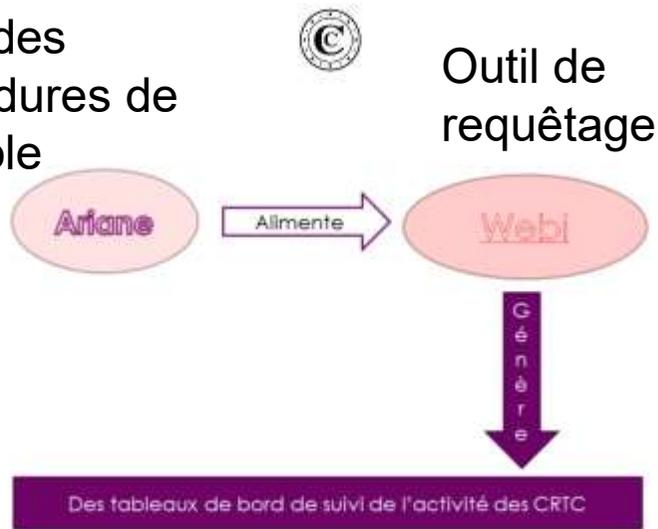
Les **objectifs** :

- L'appui aux missions traditionnelles ;
- L'identification des risques financiers, juridiques voire pénaux ;
- l'évaluation quantitative des politiques publiques ;

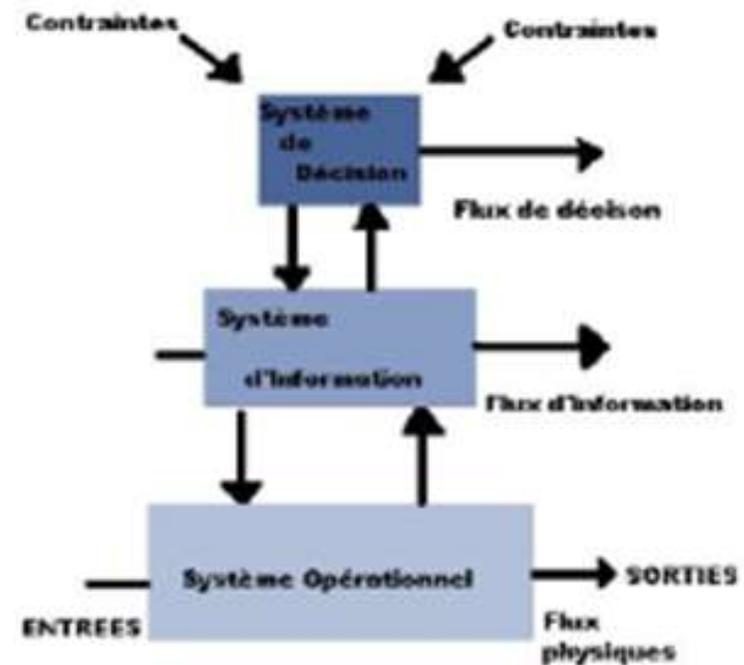


0.3 – DU SUIVI DE L'ACTIVITÉ AU SYSTÈME DE DÉCISION

Suivi des
procédures de
contrôle



Tableaux de bord de suivi de l'activité





0.4 - POURQUOI S'INTÉRESSER A CES DONNÉES INFORMATIQUES ?

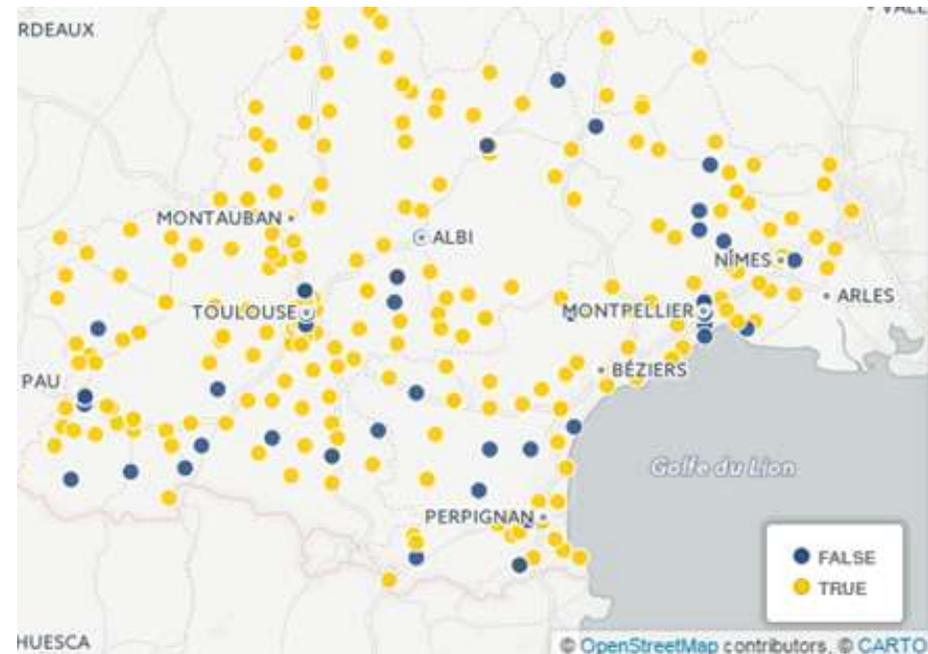
- Une disponibilité plus rapide voire immédiate de l'information : d'où une réduction des délais de publication des rapports (nowcasting),
- Une information à une échelle plus fine : Les systèmes d'information permettent désormais de produire des analyses localisées, à l'euro près ;
- Des mesures objectives, robustes face à la contradiction du contrôlé ;
- Des gains de productivité : une réduction de la charge des enquêtes ;



0.5 – VERS UN SYSTÈME D'INFORMATION DÉCISIONNEL ET GÉOGRAPHIQUE ?

Sans de nouveau développement informatique, sur la base de ces anciens outils de suivi de l'activité, il peut être souligné que l'analyse de la couverture du territoire peut utilement s'appuyer sur des outils cartographiques

Identification en Occitanie des trésoreries dont aucun organisme n'a fait l'objet d'un contrôle sur la période 2006- 2016





1^{ÈRE} PARTIE

L'OUTIL D'ANALYSE FINANCIÈRE
(ANAFI)



1.1 – Le recueil et l’analyse des flux de données des comptabilités publiques

- Le projet a été initié fin 2012 en remplacement d’un outil pré-existant qui avait été confié à un prestataire extérieur et qui était devenu largement inadapté.

- Ce nouveau projet, développé en interne, est venu en soutien des équipes de contrôle des chambres dès janvier 2014. Il propose, en quelques clics et quelques secondes un accès aux données comptables (2009 et suivantes) des services publics locaux ainsi qu’une analyse financière rétrospective **automatisée** des collectivités locales.

- L’outil doit apporter des éléments de réponse aux deux questions clés de l’analyse financière rétrospective à savoir :
 - 1/ la collectivité a-t-elle réalisée son programme d’investissements sans dégrader sa situation financière ?

 - 2/ Assume t-elle l’ensemble de ses obligations ou reporte elle des charges (entretien ou renouvellement des immobilisations) ou des risques (emprunts) sur l’avenir ?



1.2 – Objectif et modalités

- Par son **déroulé pédagogique** et les indicateurs retenus, cette analyse tend à suggérer une **méthodologie de contrôle**, sans toutefois avoir vocation à arrêter une norme opposable :
 - présenter une approche synthétique pour s'assurer rapidement du respect des grands équilibres,
 - produire une trame d'analyse homogène pour l'ensemble des collectivités,
 - garantir la cohérence avec les agrégats utilisés dans d'autres outils du ministère des finances.

- Le flux de données transmis par le ministère des finances depuis 2012 permet d'accéder aux **170 000 balances comptables** des collectivités locales, sous 18 nomenclatures différentes, correspondant aux différents types de collectivités et de services publics ;

- L'analyse financière automatisée est rendue possible par la confection de grilles adaptées à chaque nomenclature budgétaire et comptable.

- Pour rentrer plus dans le détail, l'analyse comporte deux grilles : les performances financières annuelles ; la situation patrimoniale ou bilancielle.



1.3 – Les limites et les développements récents

▪ Les **limites** :

- L'outil n'effectue pas d'analyse des restes à réaliser,
- Ni d'analyse qualitative de la dette (identification des emprunts toxiques),
- Il n'est pas conçu non plus pour contrôler la fiabilité des comptes même si des vérifications automatiques sont effectuées.

▪ Depuis son déploiement, le projet a bénéficié de **nombreux développements de modules**:

- d'analyse financière prospective,
- des services publics à caractère industriel ou commercial,
- des établissements publics de santé,
- des métropoles, etc... ;

▪ Il introduit un système de **notation financière qui lui est propre** :

- Cela renouvelle **la programmation par les risques** par l'exploitation systématisée des comptes de gestion : évaluation homogène du niveau de risque financier pour les communes ;
- Il permet de créer ainsi un réseau d'alerte paramétrable
- Cet outil connaît toutefois un usage encore limité (quelques contrôles par an dans quelques chambres).



2^{ÈME} PARTIE

RECHERCHER ET CONSULTER LES PIÈCES DÉMATÉRIALISÉES (PROJET ORC)



2.1 – L'appui à l'instruction ; des potentiels de transversalité

- Un projet également lancé fin 2012 ; qui s'inscrit dans un contexte d'extension de la **dématérialisation** de la chaîne comptable et d'affirmation du principe de **quérabilité** des pièces ;
 - → Ce n'est pas aux juridictions financières d'aller chercher les pièces mais aux comptables de les fournir a priori.
 - → un archivage électronique mutualisé : le silo de données commun conservant les pièces justificatives à l'appui des comptes de gestion (État et collectivités territoriales) ;

- L'outil est opérationnel depuis mars 2017 : avec les données 2013, 2014 et 2015. A terme l'outil stockera 10 exercices comptables.

- Une campagne de déploiement avec des formations de référents puis en cascade de l'ensemble des personnels + la délivrance d'habilitation d'accès.



2.2 – LA CONSTITUTION DE LIASSES ÉLECTRONIQUES

- Une des pratiques professionnelles classiques des juridictions financières était la fouille de liasses papiers : les équipes de contrôle avaient l'habitude de devoir plonger au sein de milliers de pages afin de rechercher l'information spécifique qu'elles recherchaient.



Le nouvel outil permet la recherche de l'information directement sous format électronique et l'assemblage d'un dossier, d'une liasse, paramétrée de façon ad hoc, également sous format électronique, puis de la télécharger.

Ex : recherche sur les marchés publics d'une collectivité locale ; les subventions ; les éléments de paye.

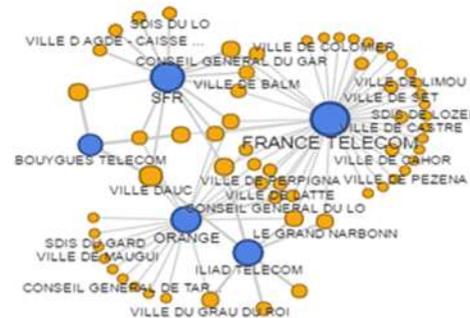


2.2 – LES RECHERCHES TRANSVERSALES

- Le principe : Les pièces comptables permettent de connaître par exemple les réseaux de tiers (fournisseurs, bénéficiaires de subventions, ...) de l'ensemble des collectivités territoriales du ressort d'une chambre et les montants associés.
- La limite technique : plafond de 1 000 résultats par requête.
- Enjeux : mieux appréhender les « éco-systèmes » locaux ;
- Mieux cibler la programmation des contrôles ;

Ex : les associations percevant le plus de subventions de fonctionnement en 2015 en Occitanie.

Carte relationnelle



| Étiquettes de lignes | Somme de Mt TTC ligne |
|----------------------------------|-----------------------|
| TAM HORODATEURS | 47 163 380 |
| TCN | 36 081 032 |
| CFTPM CORPOR FRANCAISE TRANSPORT | 22 400 191 |
| TAM | 21 422 476 |
| KEOLIS SA | 16 221 622 |
| SUD DE FRANCE DEVELOPPEMENT | 13 829 783 |
| OPERA ET ORCHESTRE NATIONAL MPT | 13 000 000 |
| SAS TRANSDEV URBAIN BMT | 10 320 082 |
| CMTE REG TOURISME MP | 10 139 150 |
| HERAULT SPORT | 7 601 505 |
| BTP CFA LR | 7 568 593 |
| TPMR TOULOUSE | 6 036 627 |
| TAM TRANSPORT AGGLOM MONTPELLIER | 5 861 553 |
| SEM TM SOCIETE D ECONOMIE | 5 346 470 |
| CROIX ROUGE FRANCAISE | 5 059 034 |
| CFTPM CORPOR.FRANCAISE.TRANSPORT | 4 854 159 |
| COSAT COMITE OEUVRE SOCIALE AG | 4 124 832 |
| HÉRAULT TOURISME | 4 124 741 |
| OPERA ET ORCHESTRE NATIONAL DE M | 4 000 000 |
| MADEELI AGENCE REG DEVPT EXPORT | 3 385 787 |



3^{ÈME} PARTIE

L'AUTOMATISATION DE LA FOUILLE DES DONNÉES MASSIVES DE PAYE (PROJET ALTAIR)



3.1 – EXPLOITER LES DONNÉES DE PAYE

- Les bulletins de paye des agents publics locaux étaient transmis sous format électronique CD-Rom aux juridictions financières lors du dépôt des comptes de gestion des comptables publics locaux.
 - Ces documents étaient toutefois peu utilisés : le ministère du budget avait développé une application d'exploration de ces données, mais l'outil était lourd à manipuler et chronophage.
 - Un développement informatique interne, sur le fondement du langage informatique « R », d'analyse statistique, a été réalisé afin de compiler automatiquement les données de paye. Il est ainsi possible de fouiller automatiquement les données des collectivités locales et des centres hospitaliers, afin de les contrôler.
-
- Objectifs :
 - L'extraction des données et le contrôle de leur fiabilité ;
 - Réaliser des tests automatisés de respect de la réglementation ;
 - Fournir un potentiel d'analyses statistiques.



3.2 – Un test transversal

- L'outil a été conçu pour aider les équipes sur 1 contrôle spécifique ;
- Mais un usage transversal peut être imaginé, sur l'ensemble des grands comptes d'une juridiction ;
- En avril 2017, un test a été mené sur les **128 grands comptes** d'Occitanie :
 - Pour tester la qualité et la complétude des données déposées ;
 - Avoir une vision d'ensemble des potentielles irrégularités;
- Cela représente **921 000** bulletins de paye contrôlés.



4^{ÈME} PARTIE

LE DIFFICILE ACCÈS AUX DONNÉES NON FINANCIÈRES



Created by ImageCatalog
from Noun Project

4.1 – L'accès au patrimoine immatériel de l'État

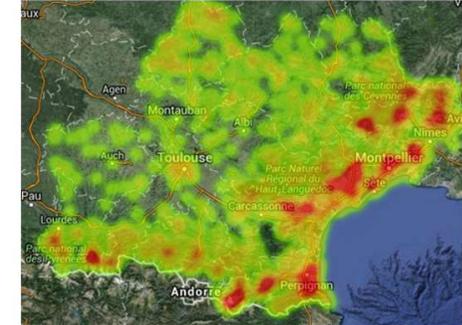
- **Enjeux** : Croiser les bases de données financières et non financières
- « **la donnée c'est du pétrole, mais avec un plus** » : c'est une matière première réutilisable. Pour découvrir cette nouvelle matière première, le croisement des bases fonctionne comme un engrenage dont les pièces, les roues dentées, vont s'imbriquer et peuvent se mouvoir.
- Pour créer une nouvelle base de données encore inexistante : exemple sur le tourisme, les piscines.
- Pour mettre en évidence des écarts à la moyenne au niveau régional.
- Pour explorer la répartition d'une valeur à l'échelle régionale.
- Confirmer ou infirmer des intuitions, les documenter et les chiffrer.
- Faire émerger des éléments d'analyse nouveaux, des schémas ou des réseaux.
- → **IMPORTANT** des relations aux services statistiques ministérielles pour :
 - identifier les bases de données issues des applications et enquêtes des services de l'Etat, et éviter le cloisonnement en silo de l'information ;
 - Avoir des échanges sur le secret statistique ;
 - Mobiliser si besoin l'Administrateur Général des données.



4.2 – exemples : les instructions en finalisation et les perspectives

- Les rapports sur le tourisme en Occitanie :
 - Répartition des produits de la taxe touristique (séjour) ;
 - Carte des retards de la gouvernance touristique ;
 - Les projets de recrutement par bassin d'emplois ;
 - Web analytics sur les sites internet de promotion touristique ;
- L'exploitation des données des marchés publics (TED) ;
- L'exploitation des données sur l'emploi dans la fonction publique locale.

Carte 1 : répartition du produit de la taxe d'habitation liée aux résidences secondaires



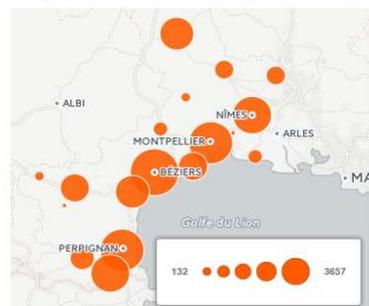
Source : données DGIEP - mise en carte avec Google Fusion Tables

Carte 2 : la carte de l'intercommunalité touristique d'après la taxe de séjour en 2014

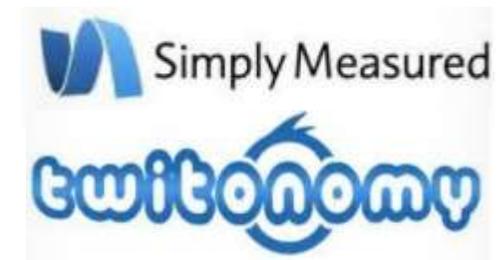


Source : CRC - anaf, d'après les comptes de gestion - mise en carte avec CartoDB

Carte 4 : les projets de recrutements en 2016 par bassin d'emploi en Languedoc-Roussillon



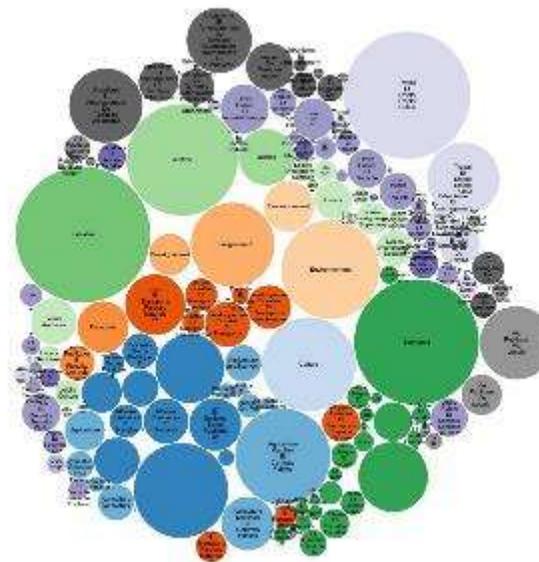
Source : enquête EMO 2016 - Pôle Emploi - mise en carte avec CartoDB





4.3 La #Datasession - juin 2017 : l'accès aux données publiques

- Apprendre des méthodes des data scientists (intégration et mise en forme des données), prototypage rapide de carte interactive, de courriers officiels.
- Tester les outils d'agilité et de brainstorming sur un sujet précis autour « d'irritants » pour le citoyen.
- Identifier les acteurs internes, locaux et nationaux qui nous aideront à progresser par la suite



LaMée Montpellier @La_... 2d

Idéation : la piscine d'images qui éveille les idées et le collectif...

#datasession @Courdescomptes @PrefetOccitanie @La_Melee





4.4 – Les limites : avantage et inconvénients de ressources centralisées

- avantages de la centralisation : la Cour des comptes s'est dotée de **data scientists** dans un pôle données national (ressource d'expertise interne forte) ;
- → mais fort risque de **congestion** : toutes les demandes convergent vers une équipe limitée ;
- → importance de diffuser la « culture de la donnée »/**data literacy** au sein de l'ensemble des équipes.
- → développer des **formations** qui vont au delà que l'initiation/expertise sous Excel.
- Exploiter les développements de l'open data : Data.gouv.fr / Insee.fr/<http://webstat.banque-france.fr/fr>
- Ex : Les recherches avancées sous Google : Combiner les opérateurs de recherche : site: .fr filetype : xls culture
- Ex : la programmation informatique, le « web scraping » → demain des juges/auditeurs codeurs ?

```
import pandas as pd
import os
from selenium import webdriver
from selenium.webdriver.firefox.firefox_binary import FirefoxBinary
from bs4 import BeautifulSoup
import re
import time
```

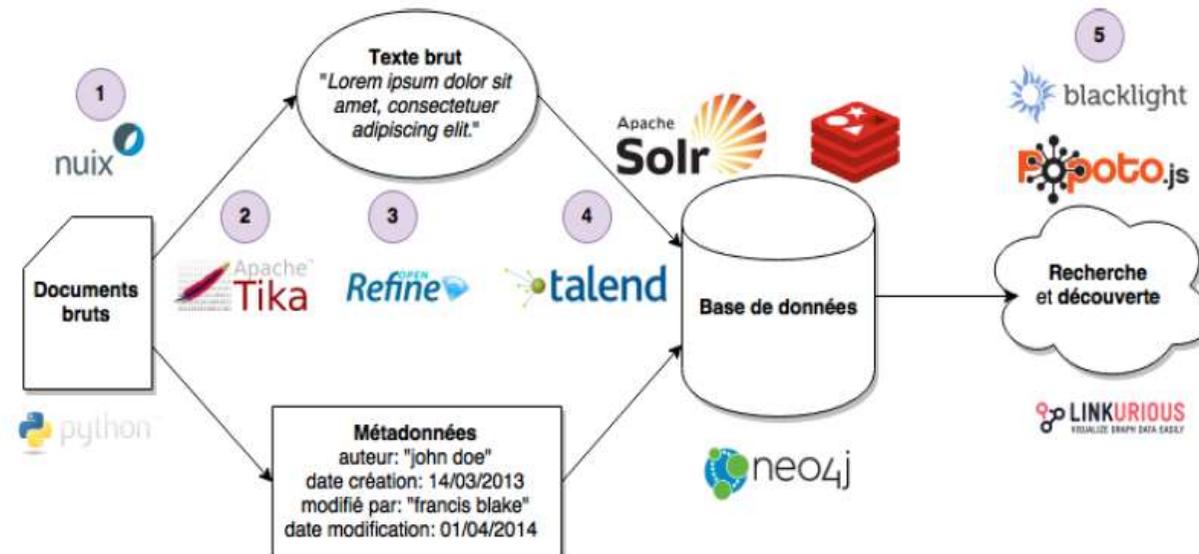
```
def scrap( input_path, output_path, binary_path, ind_debut , ind_fin ):
    start_time = time.time()
    old_index_start= 0
```



4.5 – retour d'expérience des data journalistes : les #panama papers

- 11,5 million de documents (dont 4,8 million d'emails) ;
- Des données hétérogènes et non structurées :
 - .doc, .pdf, .xls, .msg, images.
 - Essentiellement du texte (peu de chiffres)
- Des données complémentaires en Open Data :
 - Registre de commerce luxembourgeois
 - Société.com, Verif.com

Chaîne de traitement de l'information des Panama Papers



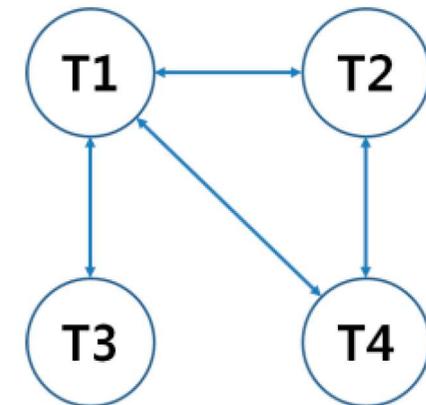


3.3 – Imaginer des usages avancés : exemple des enjeux de probité

- **A partir de la base des tiers en relation avec les collectivités du ressort il est possible de croiser les données et de détecter des enjeux de probité (prévention des conflits d'intérêt).**

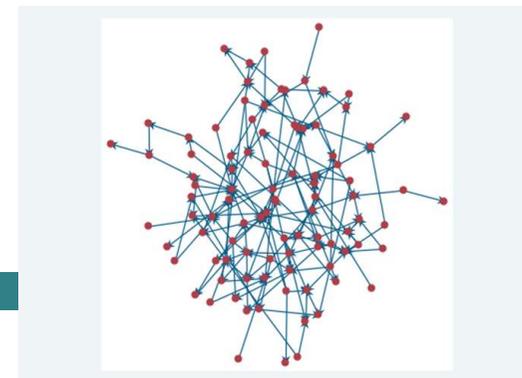
Exemple : Les personnes élus (T2) d'une collectivité (T1) qui subventionne une association (T4) où ils sont membres ;

→ s'appuyer ensuite sur les pièces dématérialisées (délibération) pour l'instruction ;



Autre exemple : les élus d'une collectivité qui a des relations financières avec une entreprise prestataire dont ils sont gérants/actionnaires/... .

- Nécessiter de tester sur des projets concrets pour évaluer le potentiel de ces méthodes, des sources : recourir aux sandbox





4.1 – Raconter des histoires (la culture de l'image)

- analyser les contenus (structures et tendances) et les faire mieux partager
- Premier principe de la visualisation des données : faire simple pour diminuer le fossé entre les chiffres et le lecteur.

4.2 – Ouvrir les données associées au rapport (participer à l'Open data)

- Depuis la data session de 2016 la Cour des comptes publie les données associées à ses rapports. La chambre Occitanie, va le faire, pour la première fois localement en 2017

4.3 – évaluer son impact (web analytics)

- compléter les indicateurs de retombées de la presse papier ;
- S'appuyer sur les outils en ligne et les relais d'influence ?