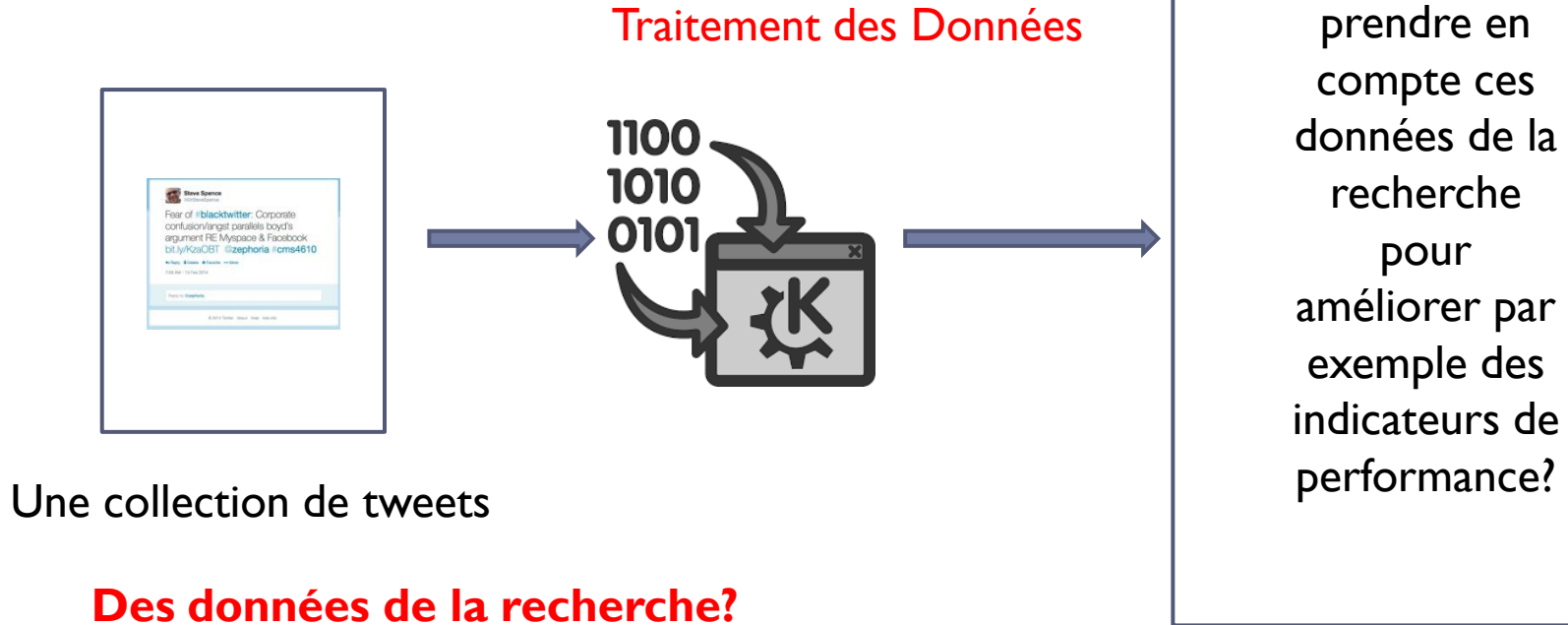


# **A particular research data: a set of tweets**

Marie-Noelle BESSAGNET  
Laboratoire Informatique LIUPPA  
PAU (France)

# Objectif de ce travail de recherche



# Les 5 dimensions

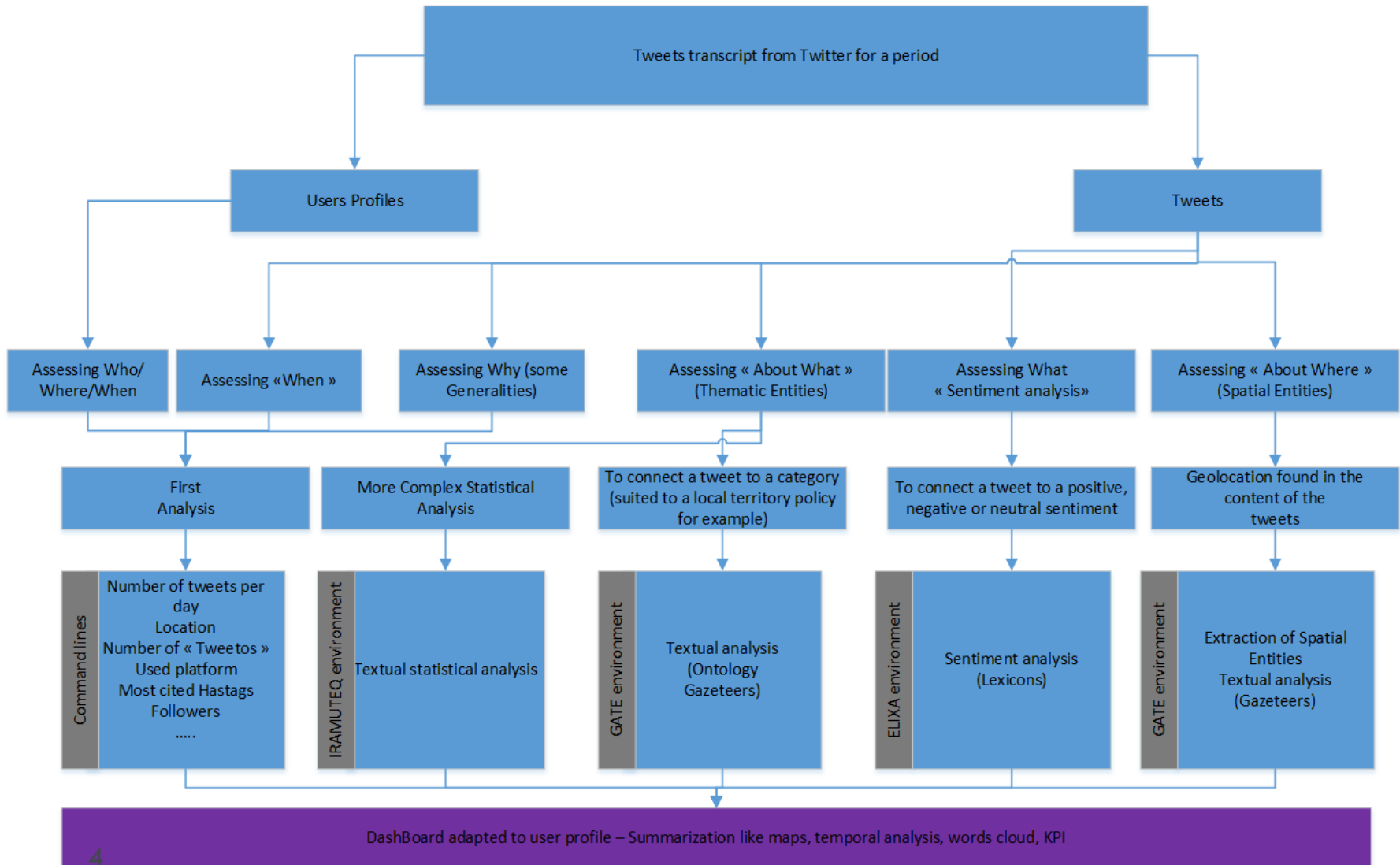
---

Who  
Where  
When  
What  
Why

**Cinq questions importantes sur  
le contenu des données de la recherche**

Afin d'aider les chercheurs en SHS à analyser leurs données de la recherche

# Un framework générique



# Deux cas d'étude

---

- ▶ **Cas d'étude créés avec des collègues enseignants-chercheurs en gestion**
- ▶ **Cas d'étude 1 : Un ensemble de tweets liés au hashtag « Bordeaux Metropole »**
- ▶ **Cas d'étude 2 : Un ensemble de tweets liés au « Bearn »**

# Deux cas d'étude

---

- ▶ **Cas d'étude I : Un ensemble de tweets liés au hashtag « Bordeaux Metropole »**
- ▶ 897 tweets collectés du 15 Janvier 2017 au 25 Janvier 2017.
- ▶ 4 caractéristiques nous intéressent dans chaque tweet :
  - ▶ Est ce que le tweet est bien lié au sujet “Bordeaux Metropole” ?
  - ▶ Quelle est l'opinion émise?
  - ▶ A quelle catégorie appartient le tweet?
  - ▶ Où est localisé l'émetteur du tweet?

# Cas d'étude 1 : les données à extraire

---

	Tweet place
THEMATIC INFORMATION	Category
	Opinion
SPATIAL INFORMATION	Authors city geocodes
	Tweet place geocode
TEMPORAL INFORMATION	Day and time of publication

Modèle synthétique des données à indexer

Pour les catégories, une ontologie est en cours de création : catégorie, sous-catégorie, ...

# Cas d'étude 1 : des analyses

Cas traité manuellement

Informatisation prévue dans un appel à projets...encore en attente

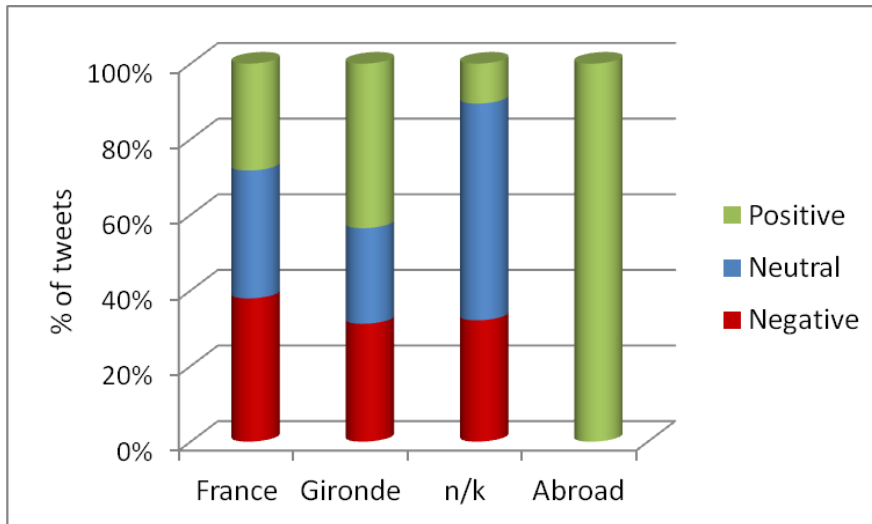
TWEETS	Id	OPINION	Category
“CE SONT LES ENTREPRENEURS QUI PERMETTENT LE RAYONNEMENT DE LA MÉTROPOLE DE BORDEAUX” @VIRGINIECALMELS <a href="https://t.co/1S9HRM8YDX">HTTPS://T.CO/1S9HR M8YDX</a>	1	Pos	Economy
UNE INFIRMIÈRE LIBÉRALE A PASSÉ SA DEUXIÈME NUIT DEVANT LA CPAM - FRANCE 3 AQUITAINE <a href="https://t.co/T8V3HV8Cw9G">HTTPS://T.CO/T8V3HV 8 Cw9G</a>	2	NEG	Economy
BORDEAUX	3		

*Exemples de tweet liés à la  
catégorie Economie*



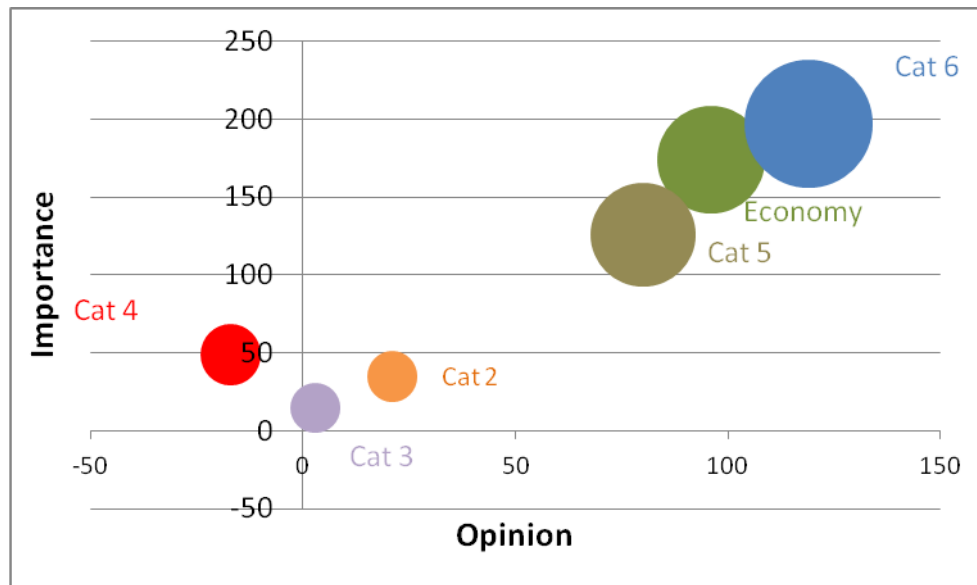
# Cas d'étude 1 : des analyses

---



Croisement dimensions spatiale (origine utilisateur) et thématique (opinion)

# Cas d'étude 1 : des analyses



Croisement dimensions thématique (catégorie) et thématique (opinion)

# Cas d'étude 1 : encore un gros travail à mener

---

1. Des analyses à détailler et approfondir
2. Une ontologie à créer : prise en compte des concepts mais également des relations entre concepts
3. L'information spatiale à extraire des tweets : non trivial
4. Un environnement informatique à concevoir et développer
5. Des réponses à appel à projets sans succès pour le moment → projet en stand-by

# Deux cas d'étude

---

- ▶ **Cas d'étude 2 : Un ensemble de tweets liés au hashtag «Bearn»**
- ▶ **7400 tweets collectés** du 13 Janvier 2017 au 31 Juillet 2017.

# Cas d'étude 2 : les données à extraire

---

	Tweet place
THEMATIC INFORMATION	Category
	Opinion
SPATIAL INFORMATION	Authors city geocodes
	Tweet place geocode
TEMPORAL INFORMATION	Day and time of publication

Modèle synthétique des données à indexer

# Cas d'étude 2 : des analyses

---

1. Cas traité informatiquement sur certains aspects
2. Evolution « territoire chaud » → marketing territorial
3. Des premières analyses sur quelques dimensions who, when, where, what
4. Un premier environnement



# Cas d'étude 2 : des analyses

---

*Les hashtags les plus cités*

*Analyse sur les hashtags géographiques*

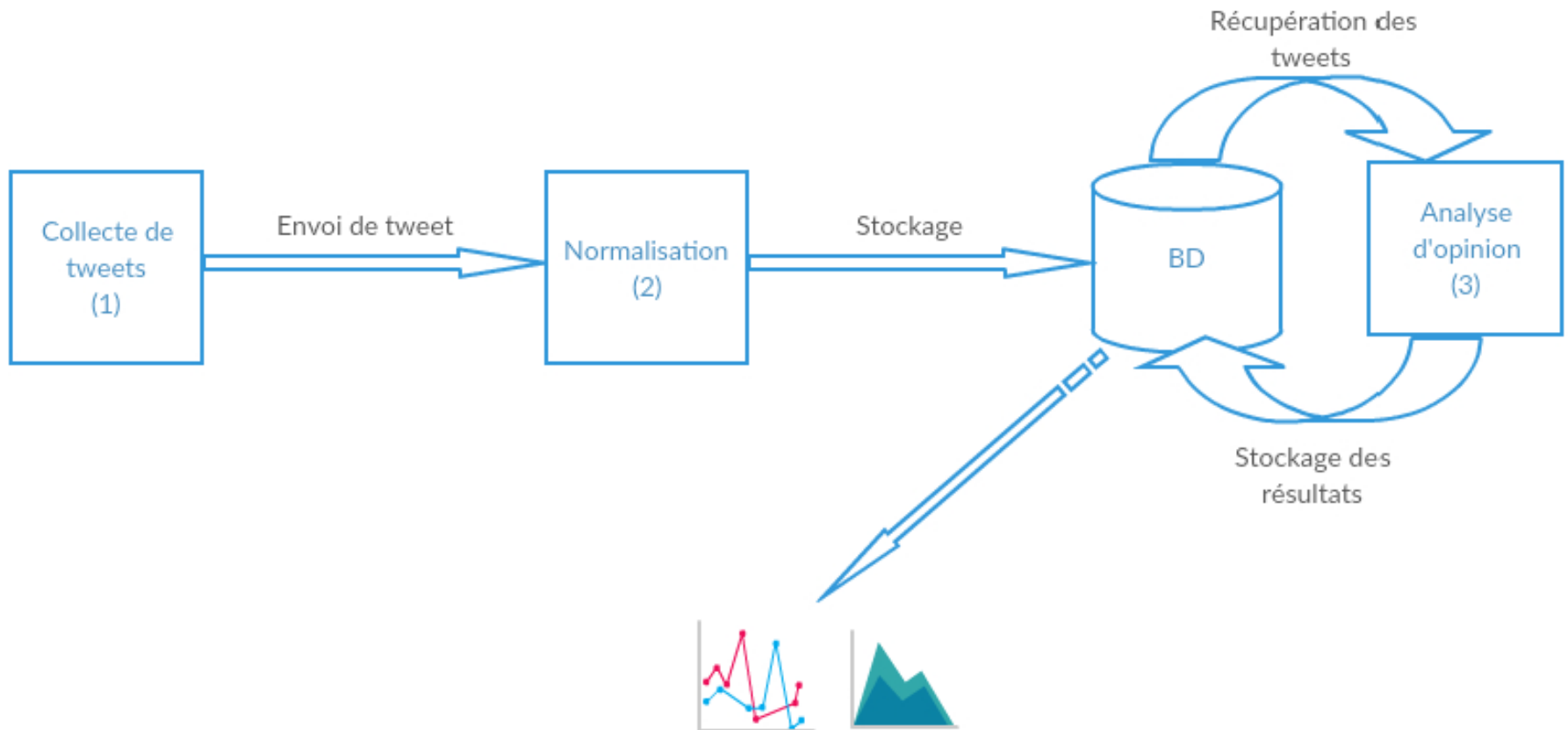
918 "biarn"  
545 "bearn"  
412 "pau"  
278 "béarn"  
257 "pyrénéesatlantiques"  
214 "pyrénées"  
150 "paysbasque"  
125 "ossau"  
115 "tourismepau"  
93 "agenda"  
91 "biarntoustèm"  
81 "tourisme"  
78 "legendadujour"  
78 "biarntoustem"  
71 "pyrenees"  
66 "oloron"  
64 "biarritz"

63 "upr"  
63 "fa2017"  
63 "bayonne"  
63 "asselineau2017"  
63 "anglet"  
59 "neige"  
59 "montagne"  
58 "photodujour"  
57 "enmarche"  
55 "magnifiquefrance"  
54 "grippeaviaire"  
46 "carattour2017"  
45 "orthez"

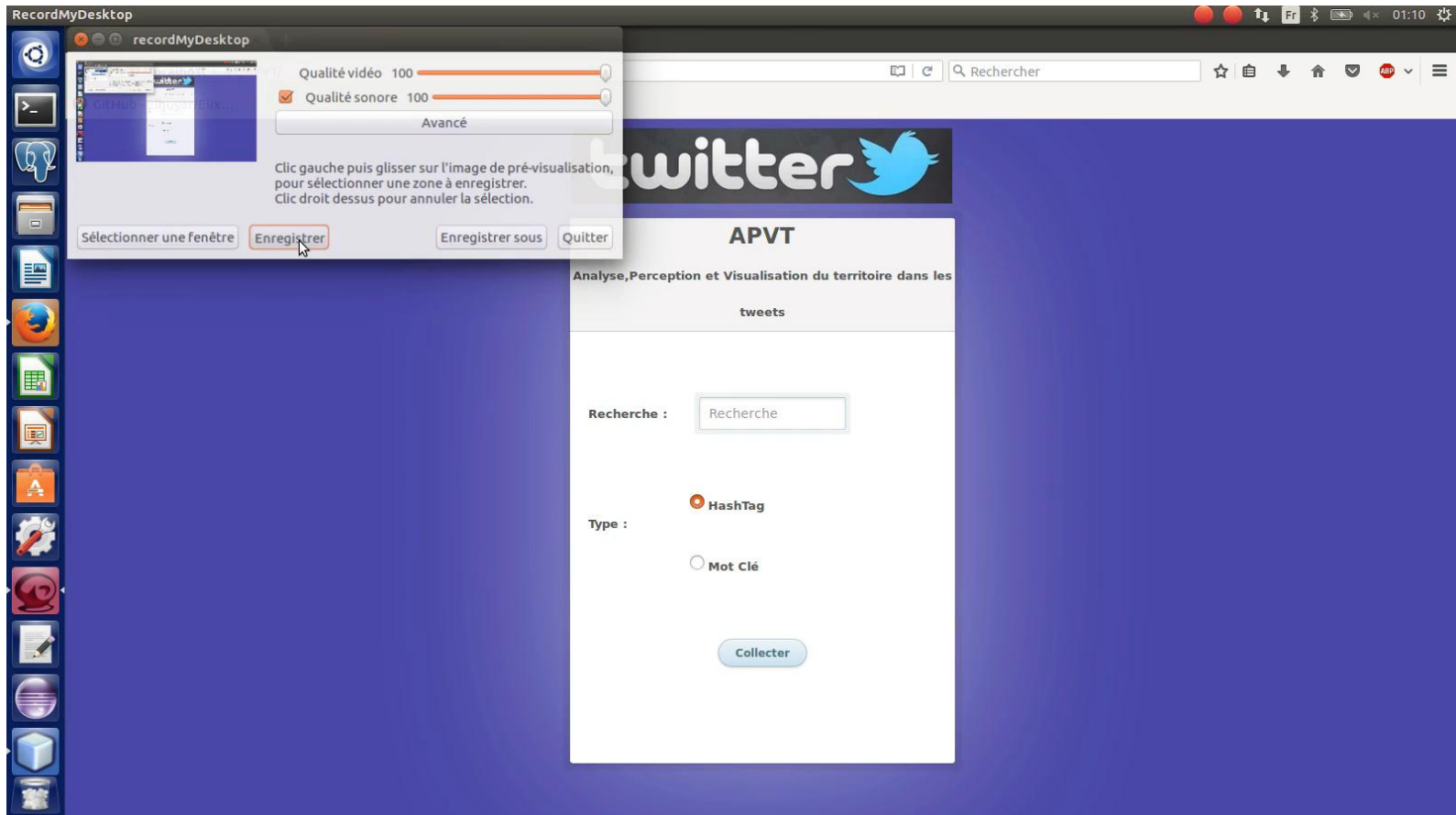




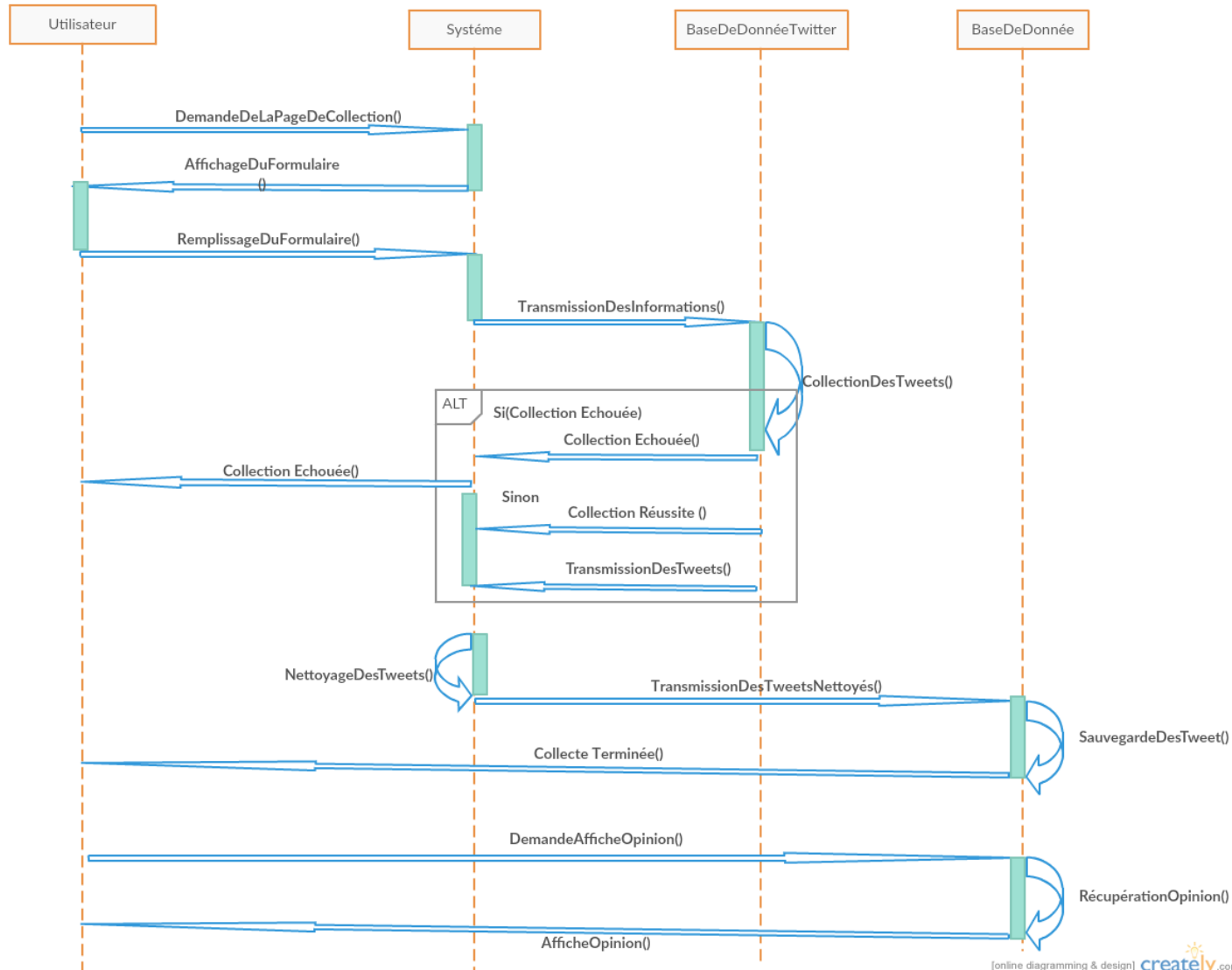
# Un premier environnement



# Un premier environnement



# Interactions Utilisateur-Systeme



# Un premier environnement

## APVT

Analyse, Perception et Visualisation du territoire dans les tweets

Rechercher :

Pseudo	Date De Publication	Tweet	RT	Localisation	Opinion
	2017	et Patricia Carricaburu!!?			
MaisonMalnou	Mon Aug 28 13:58:11 CEST 2017	Belle récolte ce matin 271kg ? #awesome #nature #chili #happy #recolte #bearn #igers #mm #work? <a href="https://t.co/HJ9cP5AtmE">https://t.co/HJ9cP5AtmE</a>	NON	Poey-de-Lescar, France	positive
radio_sudouest	Mon Aug 28 10:40:27 CEST 2017	1ère webradio destinée à la #musique et la #culture du #sudouest #paysbasque #occitanie #béarn #pyrénées #landes <a href="https://t.co/1JBQYc04Hj">https://t.co/1JBQYc04Hj</a>	NON		neutral
sabrina_glz64	Mon Aug 28 12:09:22 CEST 2017	@_melissane @seanb0y @Aliciaisbad @clay_clew @LaetitiaRebx des vôtres. À très vite à Mayol ou en Béarn. Gros bisous ?	NON	Pau, France	neutral

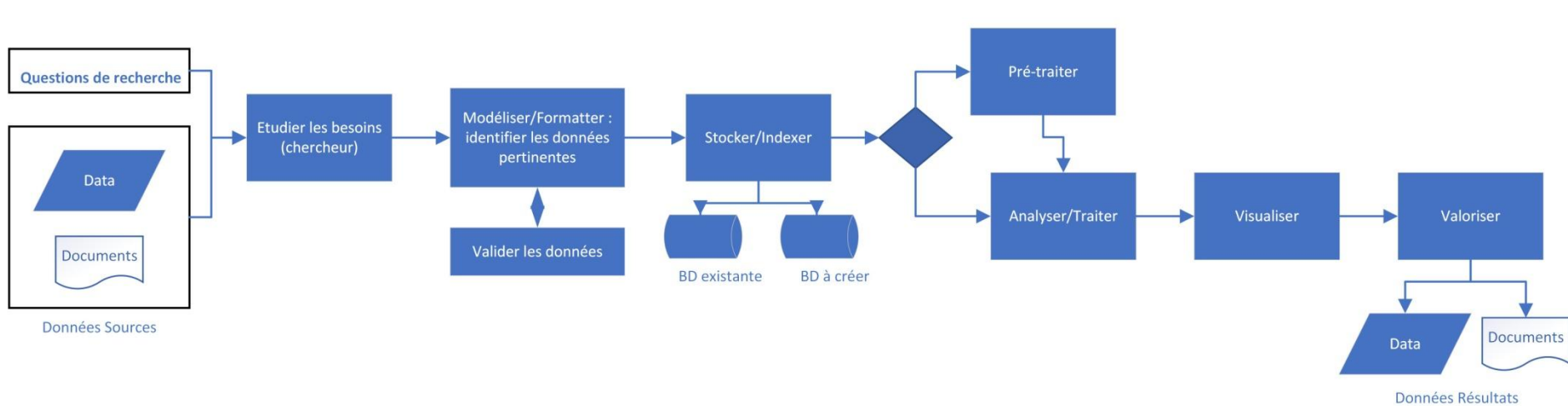


# Des Questions?

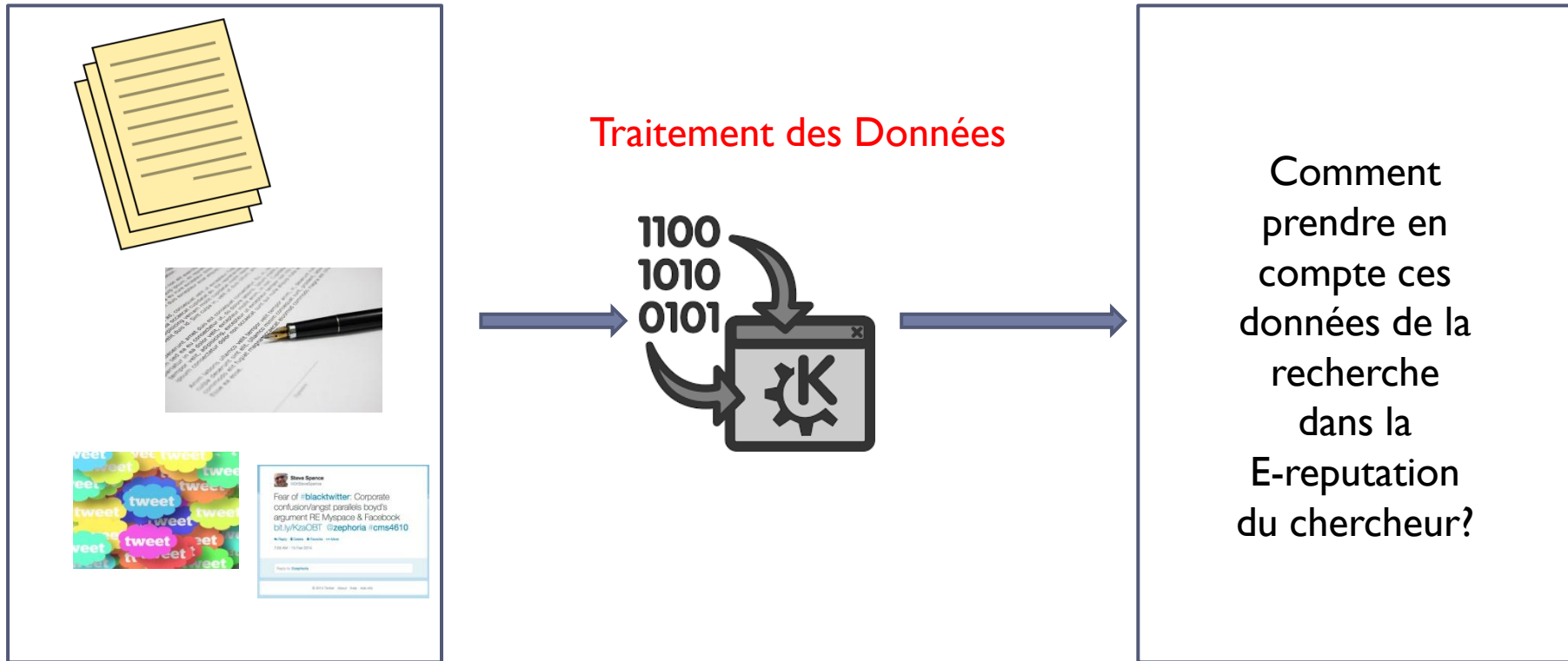
---

- ▶ Comment encourager l'adoption des bonnes pratiques en gestion de données à travers le cycle de vie des données par les chercheurs en SHS?
- ▶ Comment “donner” de la réputation à ces ensembles de données?
- ▶ Comment rendre les données de la recherche traçables par d'autres à des fins vérification et pour les rendre réutilisables (au delà du but original pour lesquelles elles ont été collectées).
- ▶ Comment concevoir une boîte à outils génériques pour les chercheurs en SHS afin de traiter et analyser leurs données de la recherche

# Prise en compte de données de la recherche plus hétérogènes – Un cycle de vie homogène



# Extension des types de données de la recherche



Des collections différentes de documents textuels non structurés

***Les données de la recherche comme une image de l'activité de recherche ou comme un indicateur de performance (Almetrics)?***

**Données de la recherche**

# Extension des types de données de la recherche

