

Webcraping : enjeux techniques et éthiques

Alexandre Cebeillac (Post-Doc UMR IDEES)

Sébastien Rey-Coyrehourcq (IGR UMR IDEES)



@reyman64

ideas

15 nov. 2019 -Paris

CIST - INFTER

Déroulement de la présentation :

Première partie

- * Généralités sur Internet
- * Principe de collecte des données

Deuxième partie

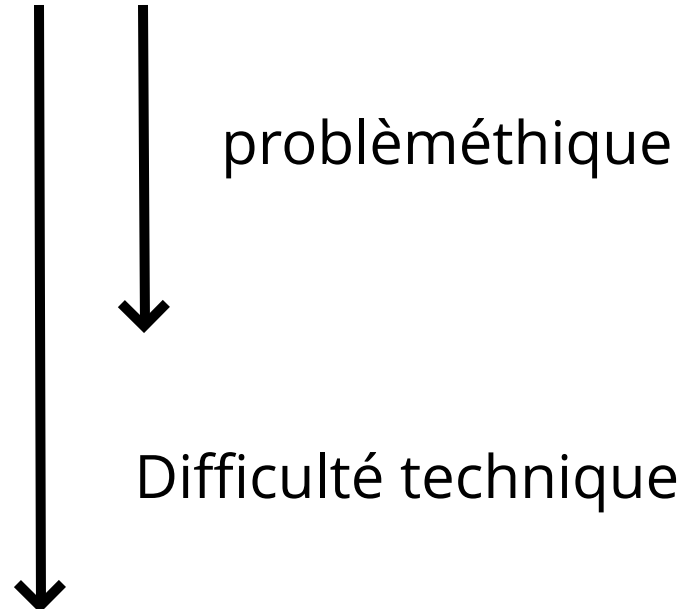
exemples d'utilisations avec contexte, méthodes, limite et application

Avec API

- * Bing
- * Flightradar
- * Airbnb
- * Twitter
- * Google

Sans API

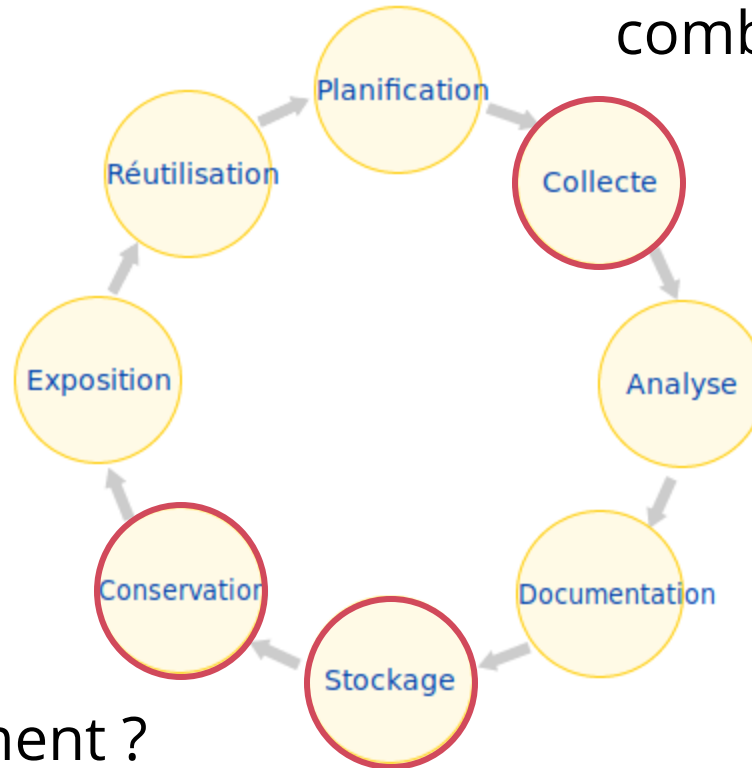
- * Géocache
- * MarineTraffic



Conclusion

Cycle des données

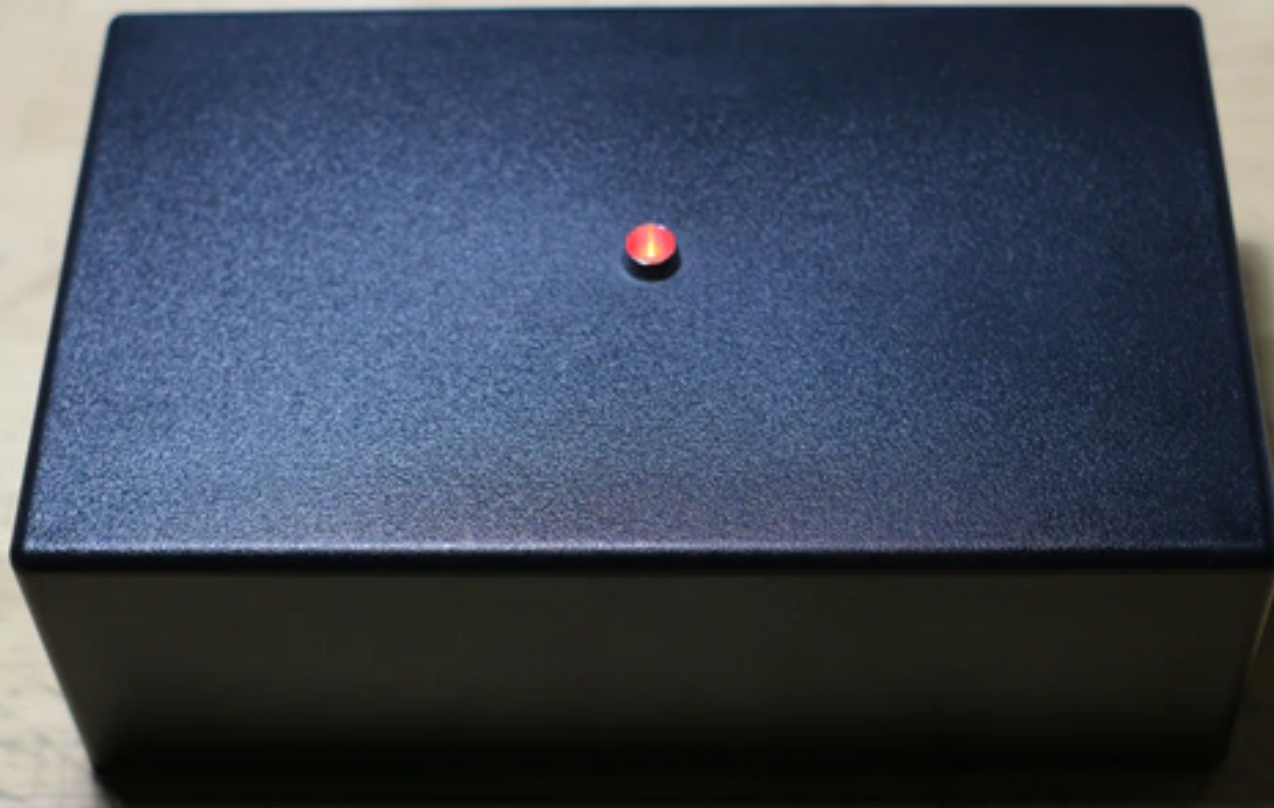
comment ?
combien de temps ?



ou ? comment ?
combien de temps ?

Src : cat.opidor

The Internet



Src : IT Crowd, S03Ep4

Internet n'est pas le Web

Internet : 1969 - 2019 (50 ans en octobre)

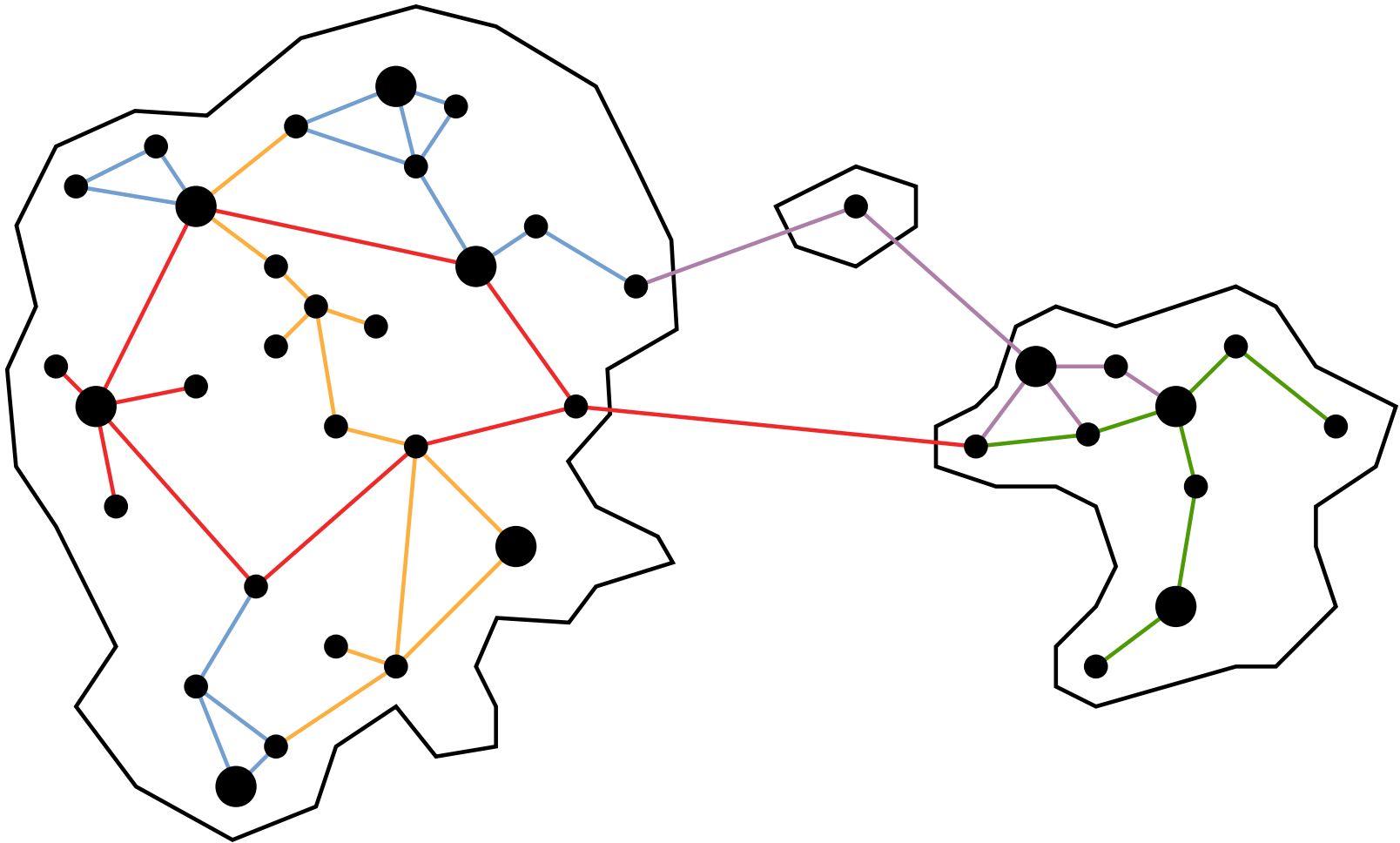
World Wide Web : 1989 - 2019 (30 ans en octobre)



Tim bernard lee (gauche), inventeur of the WWW (1989)
Vint Cerf (droite), co-inventeur du protocole Tcp/Ip

The Internet

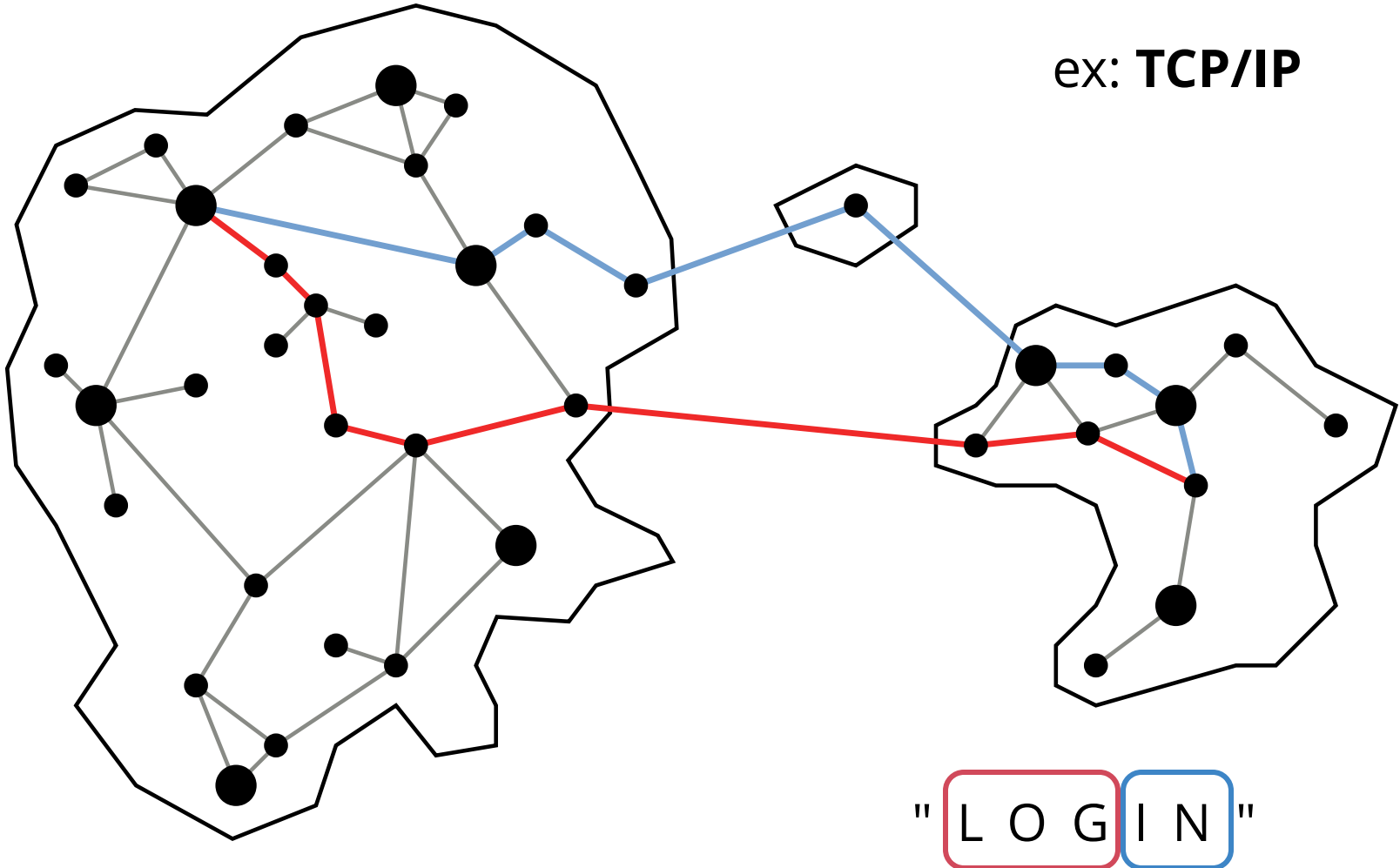
1ère couche : Infrastructures physiques



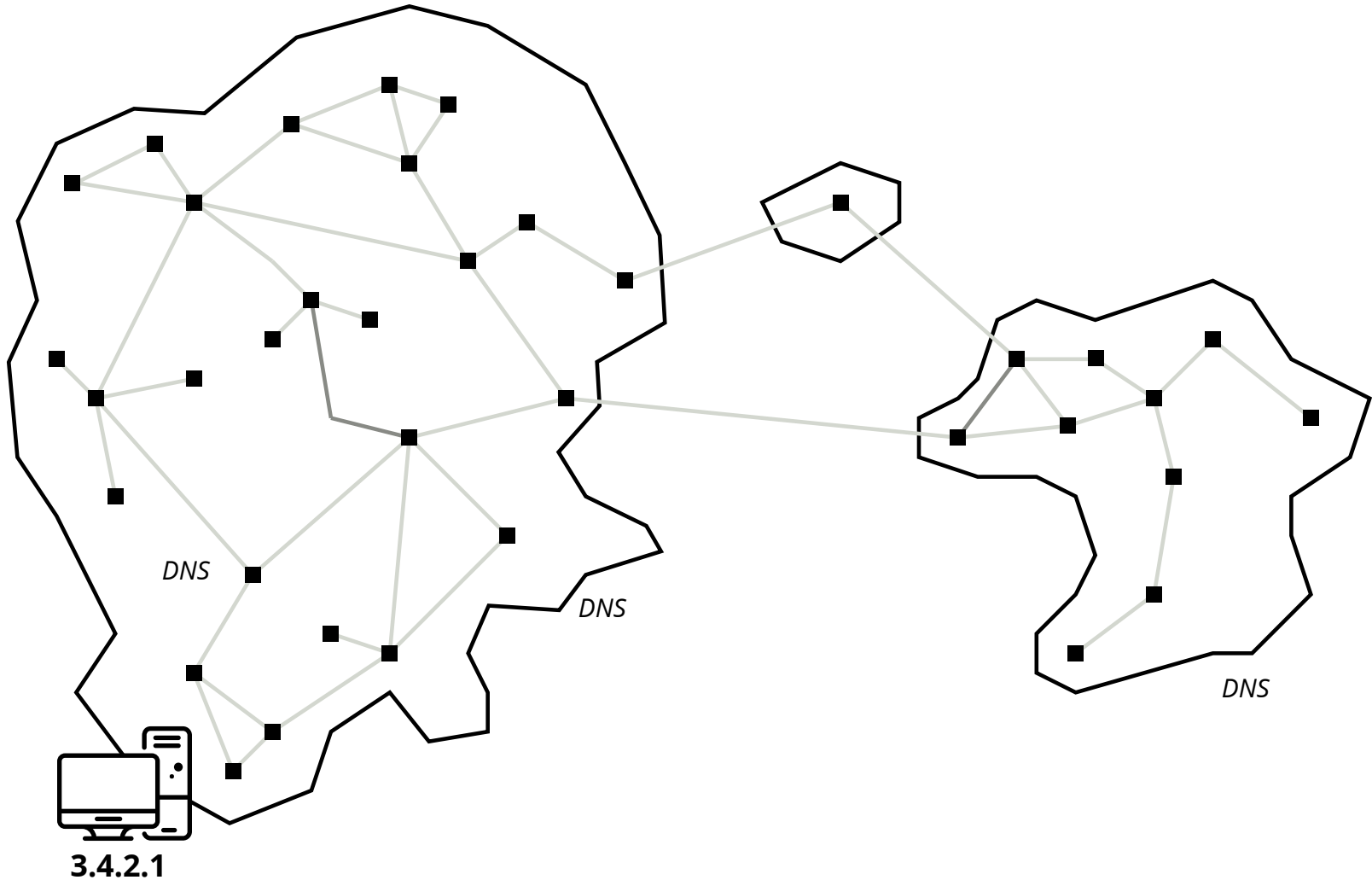
The Internet

2 ème couche : Protocoles !

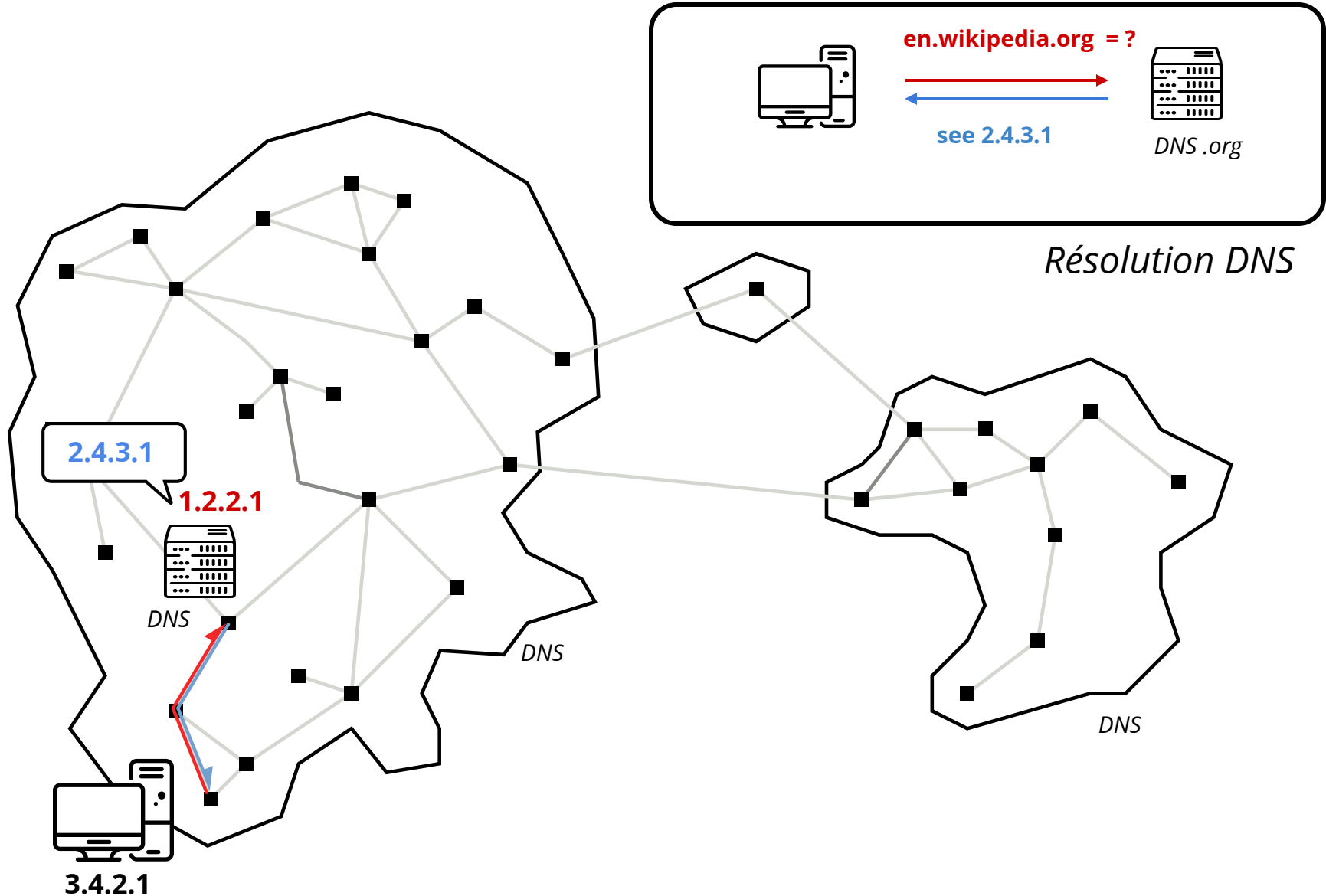
ex: **TCP/IP**



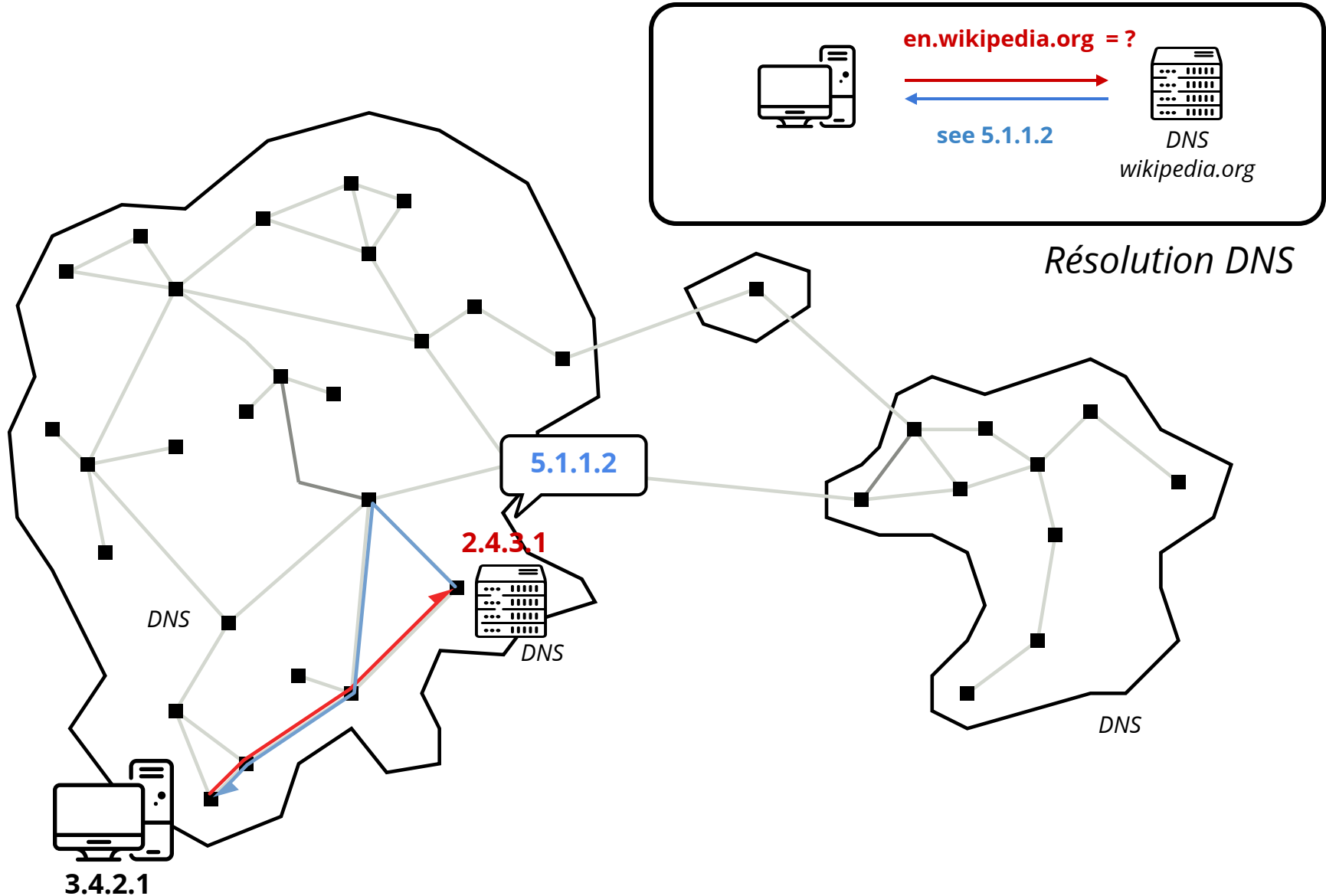
The Internet



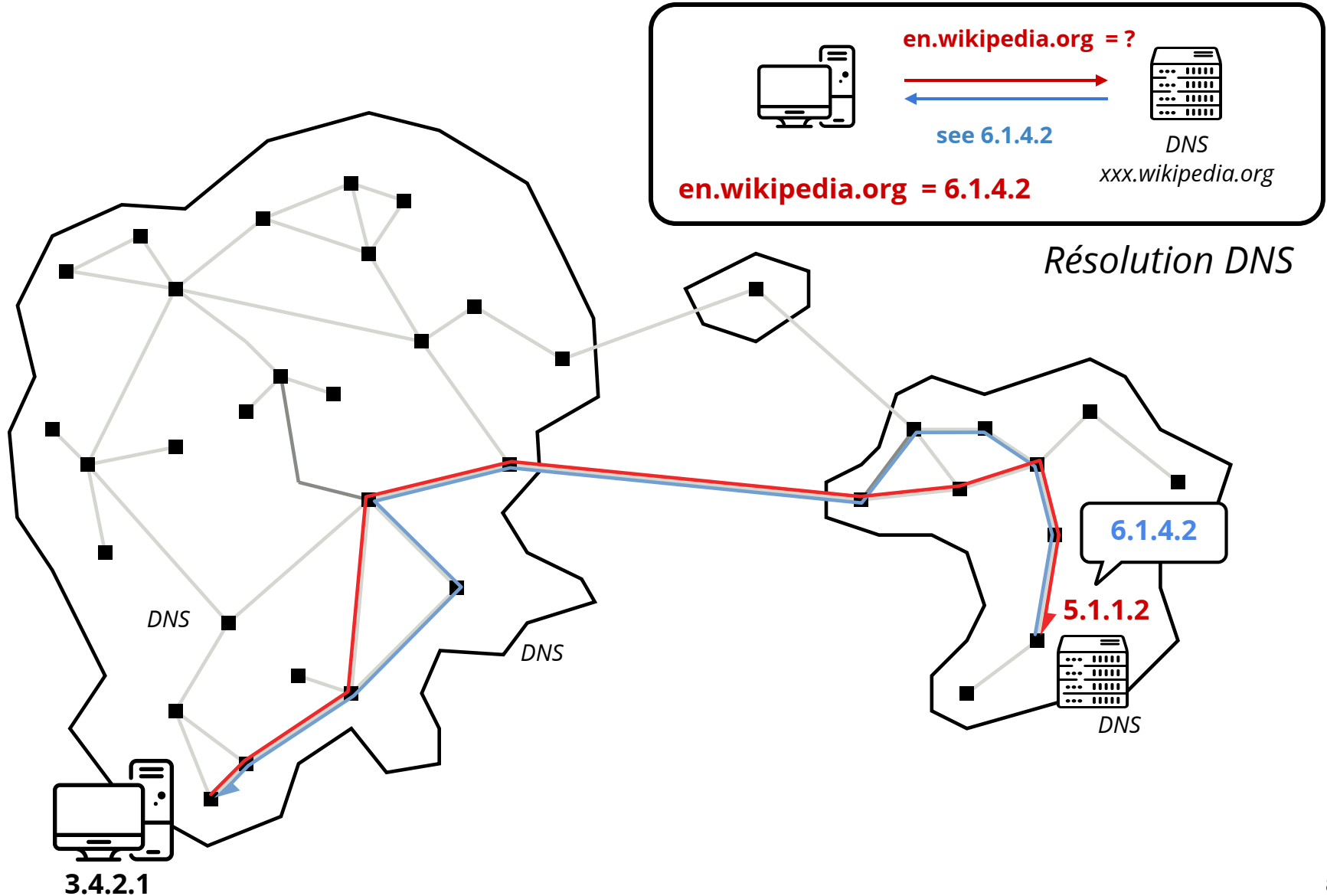
The Internet



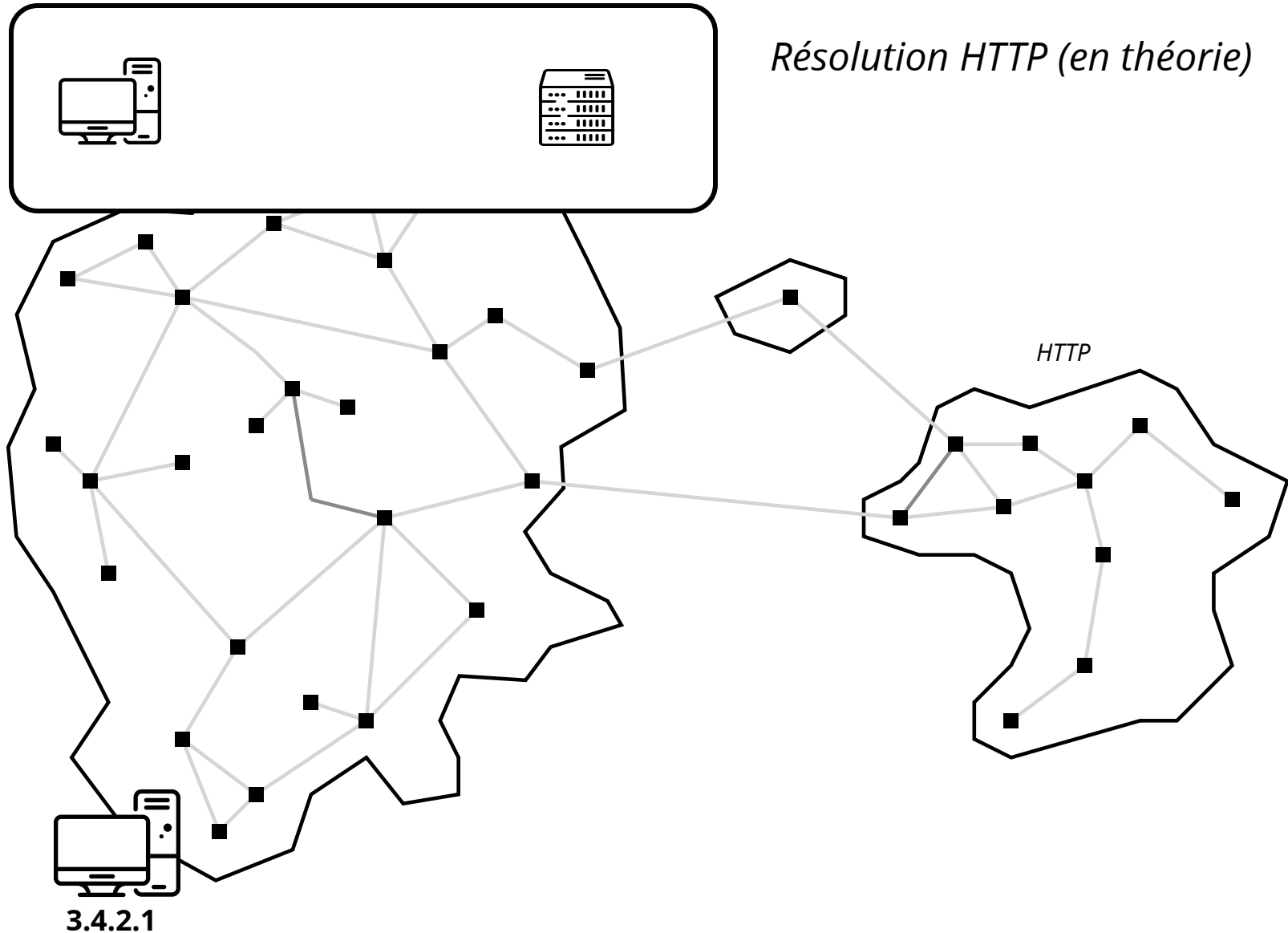
The Internet



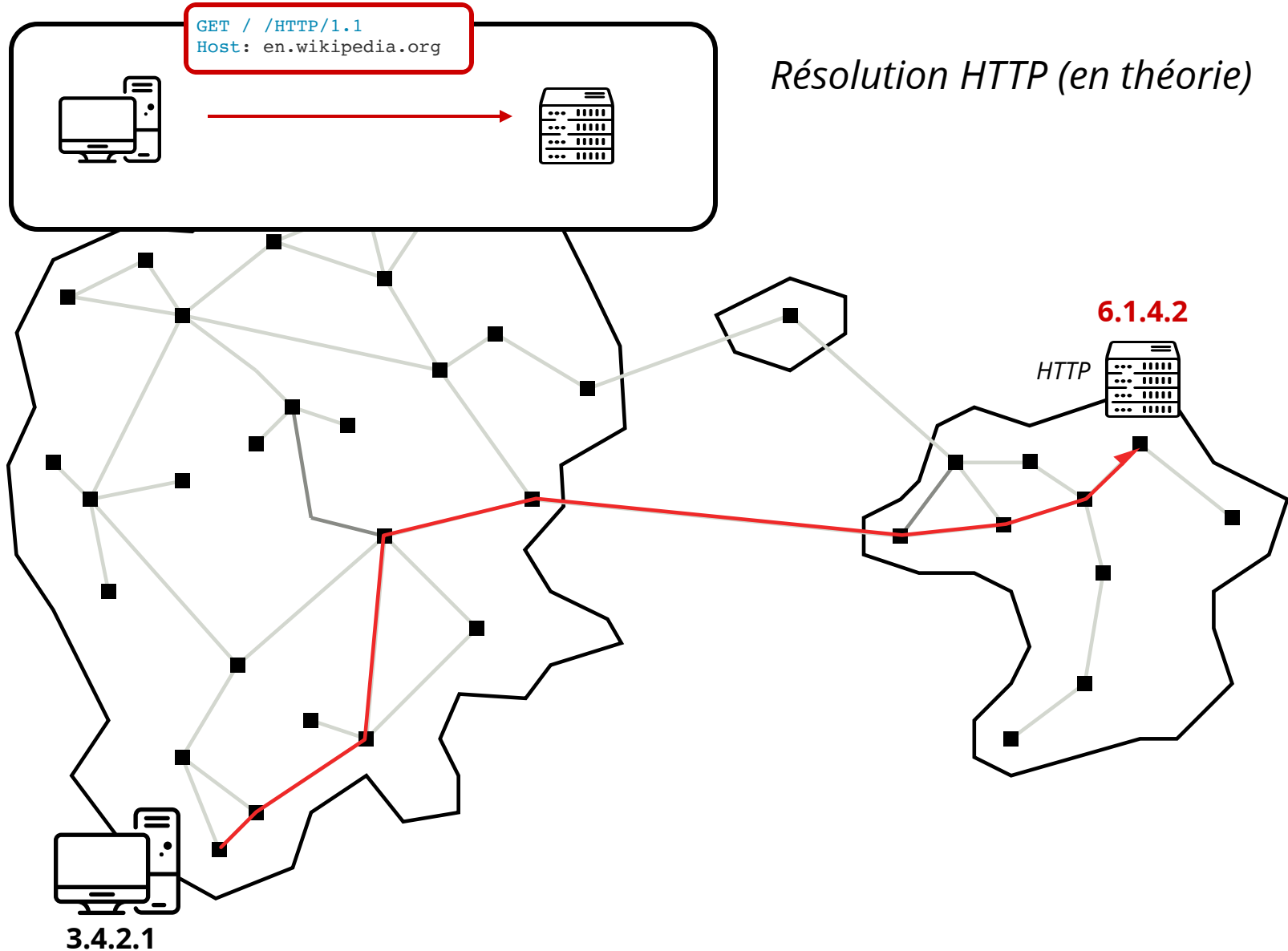
The Internet



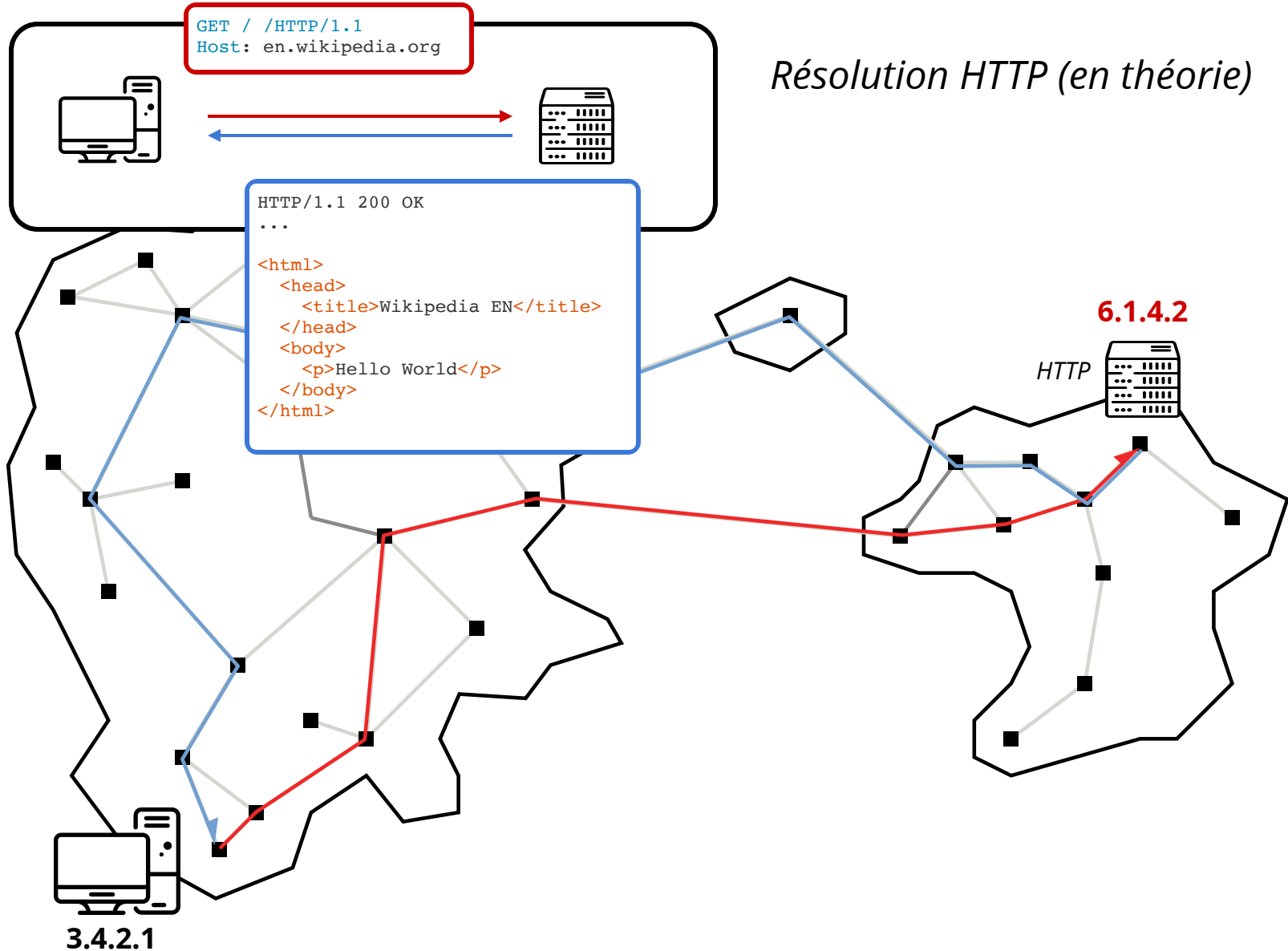
The Internet



The Internet

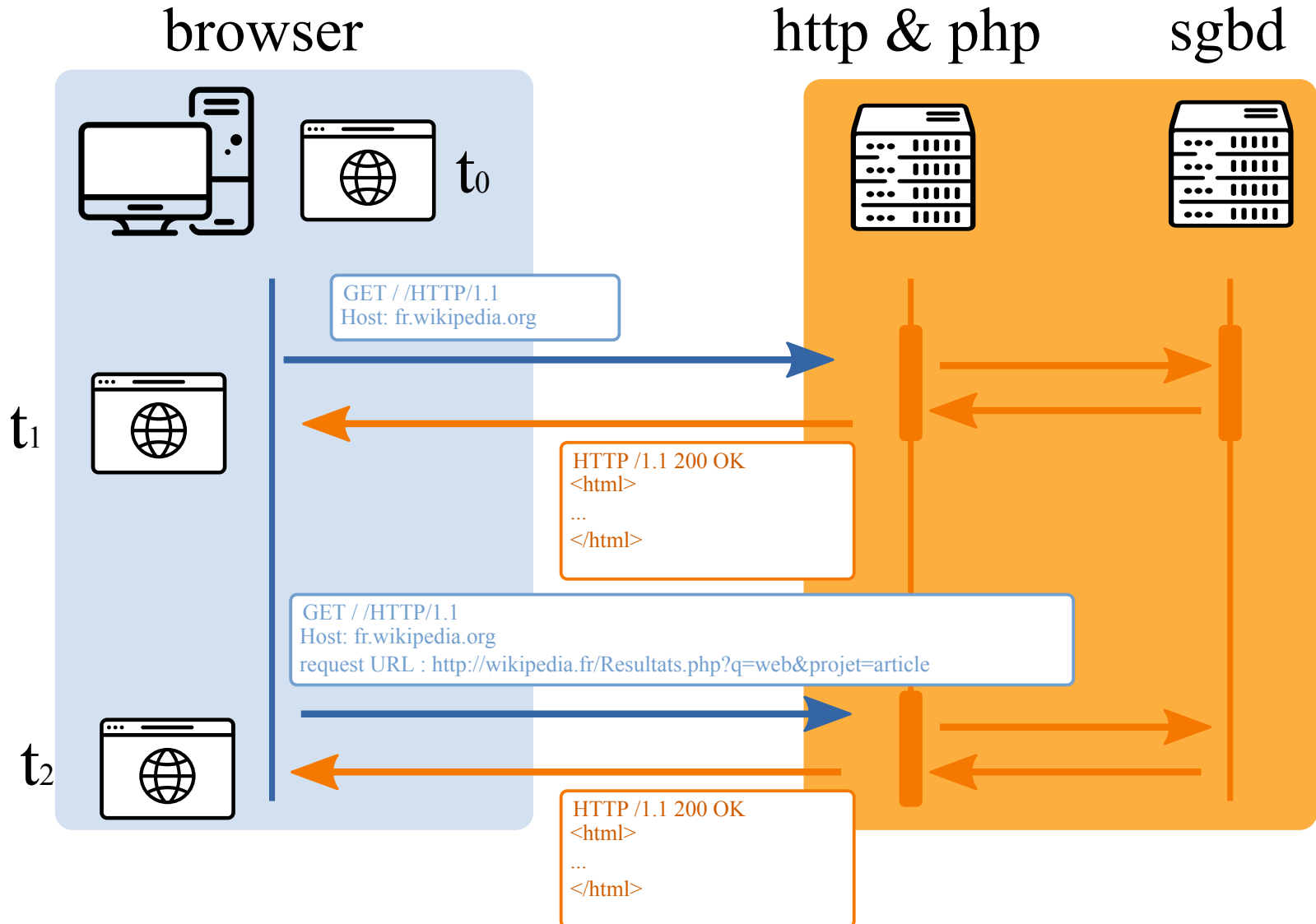


The Internet



Résolution classique

Synchrone



Page web / scraping

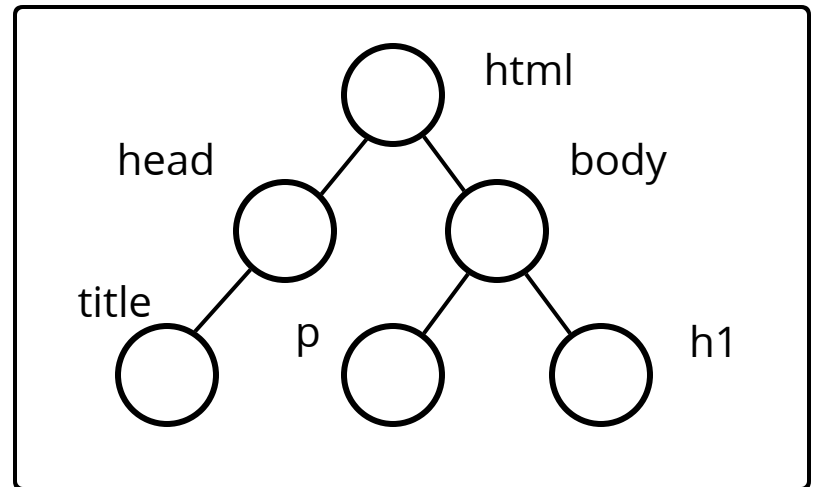
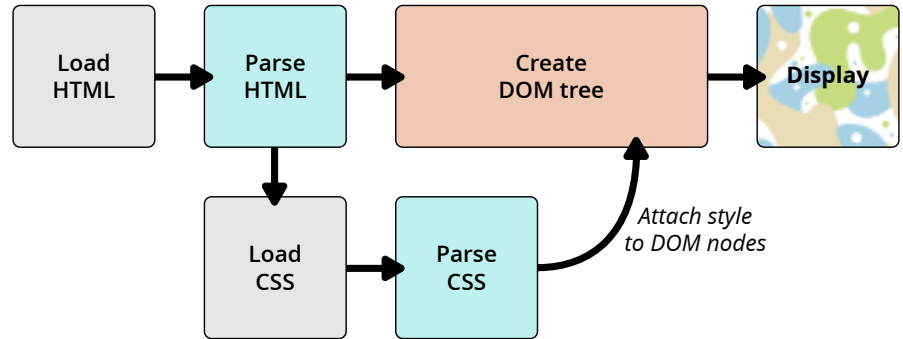
Une page HTML ?

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <link rel="stylesheet" href="css/style.css">
    <title>Test</title>
  </head>
  <body>
    <h1> Ma page web </h1>
    <p> Hello World </p>
  </body>
  <script src="js/script.js"></script>
</html>
```

```
body {
  background-color: #0080ff;
}

h1 {
  color: #fff;
  font-family: Arial, Helvetica, sans-serif;
}
```

```
let d = new Date();
document.body.innerHTML = "<h1>Today's date is " + d + "</h1>"
```

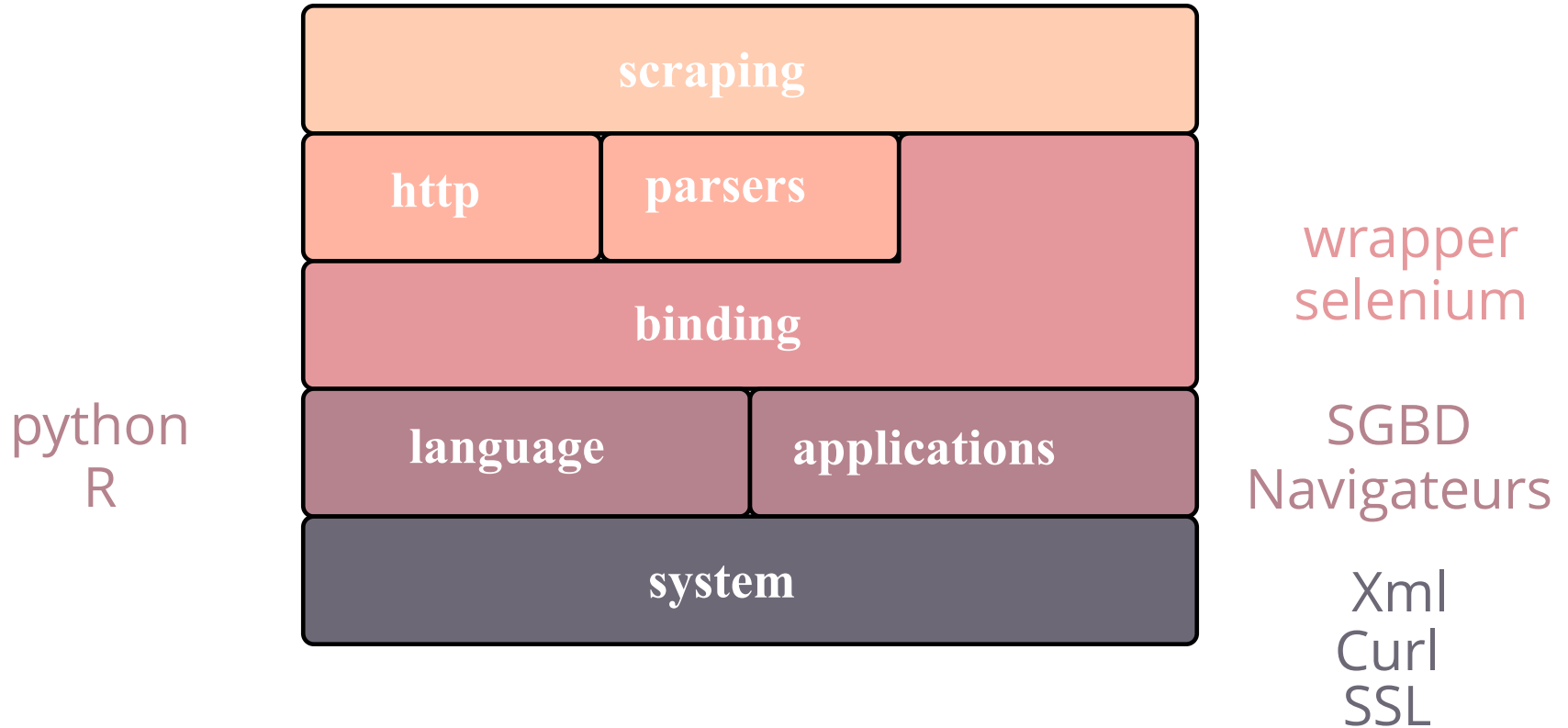


Extraction

(XPath, CSS, Regex, etc.)

Librairies/Frameworks

choisir la bonne entrée en fonction des usages

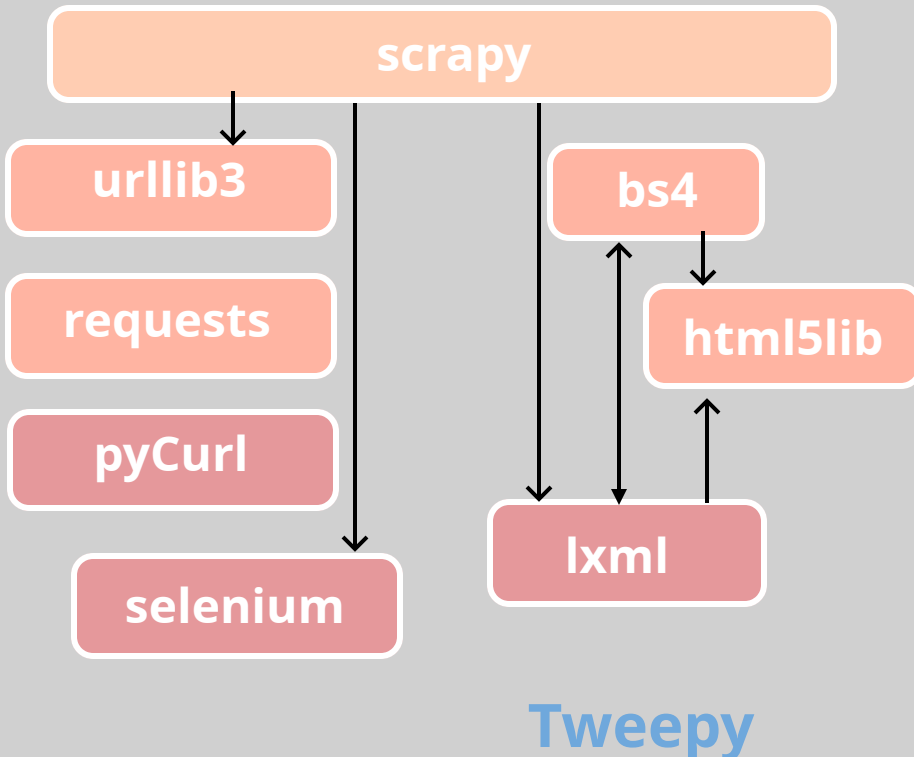


Librairies/Frameworks

choisir la bonne entrée en fonction des usages

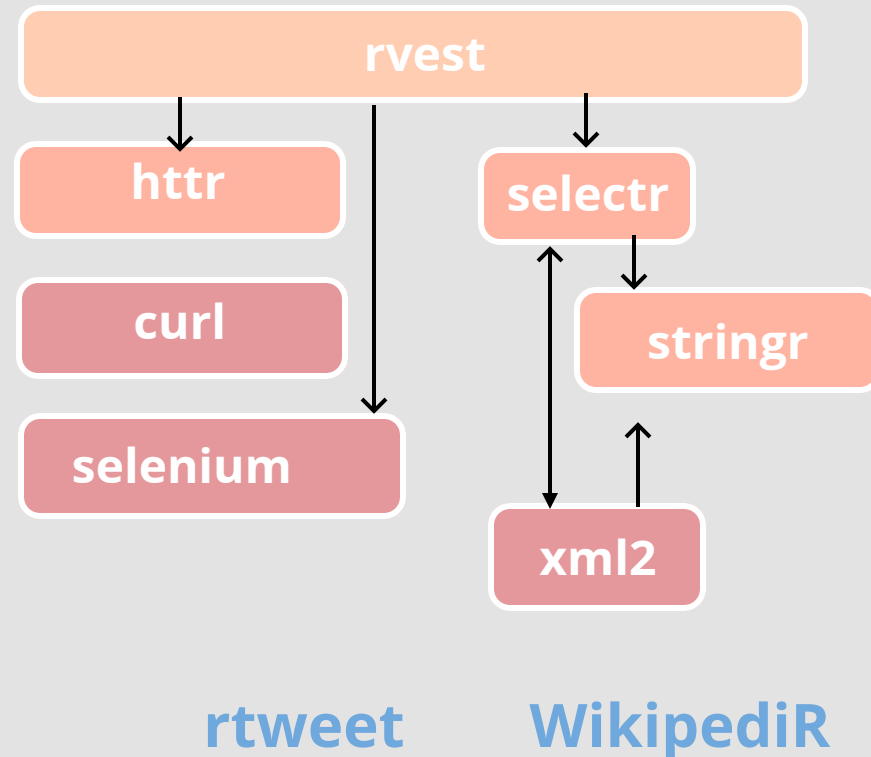
PYTHON

Collect lib. **+** Extract lib.



R

Collect lib. **+** Extract lib.



Applications

Évacuations massives



ANR Escape (2016 - 2020) / Eric Daudé

"Contribuer à la conception de systèmes d'aide à la décision dans le cas d'évacuations massives"

Un des objectifs :

Simuler les mobilités en cas de catastrophes (industrielles ou naturelles). Multisites

Comment ?

Simulation à base d'agents

Préalable :

Connaître les comportements de mobilités en condition "normale"

==> Intérêt des conditions de trafic (en plus des EMD)

Évacuations massives



ANR Escape (2016 - 2020) / Eric Daudé

"Contribuer à la conception de systèmes d'aide à la décision dans le cas d'évacuations massives"

Un des objectifs :

Simuler les mobilités en cas de catastrophes (industrielles ou naturelles). Multisites

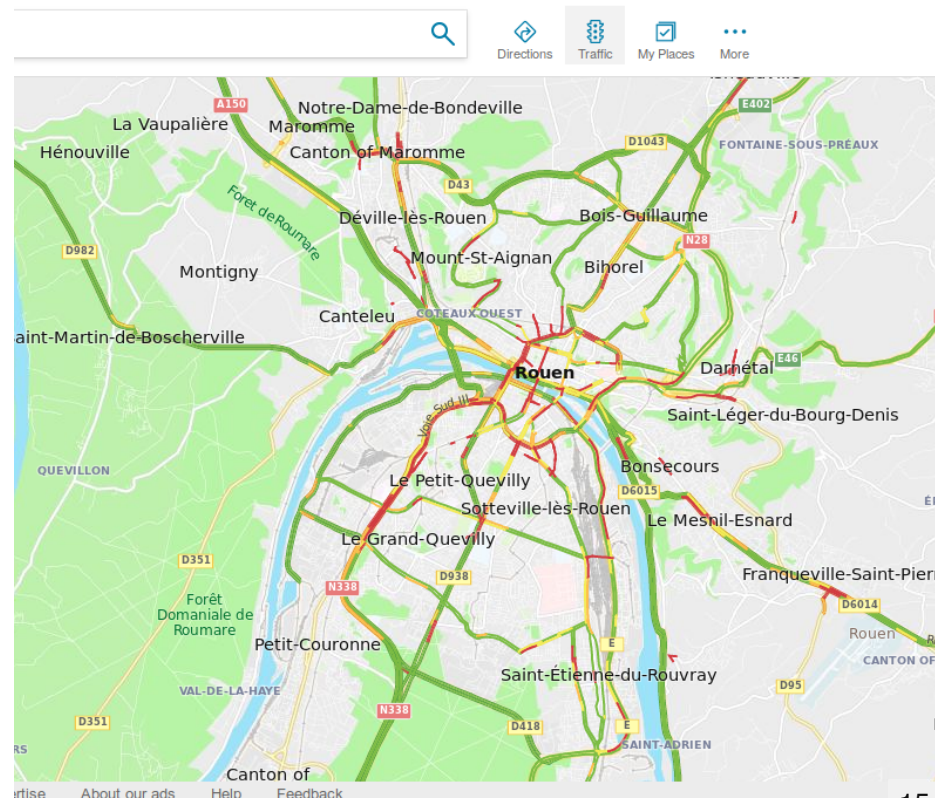
Comment ?

Simulation à base d'agents

Préalable :

Connaître les comportements de mobilités en condition "normale"
==> Intérêt des conditions de trafic
(en plus des EMD)

bing.com/maps?cp=49.4375-1.083&sty=r&lvl=12&FORM=MBEELD

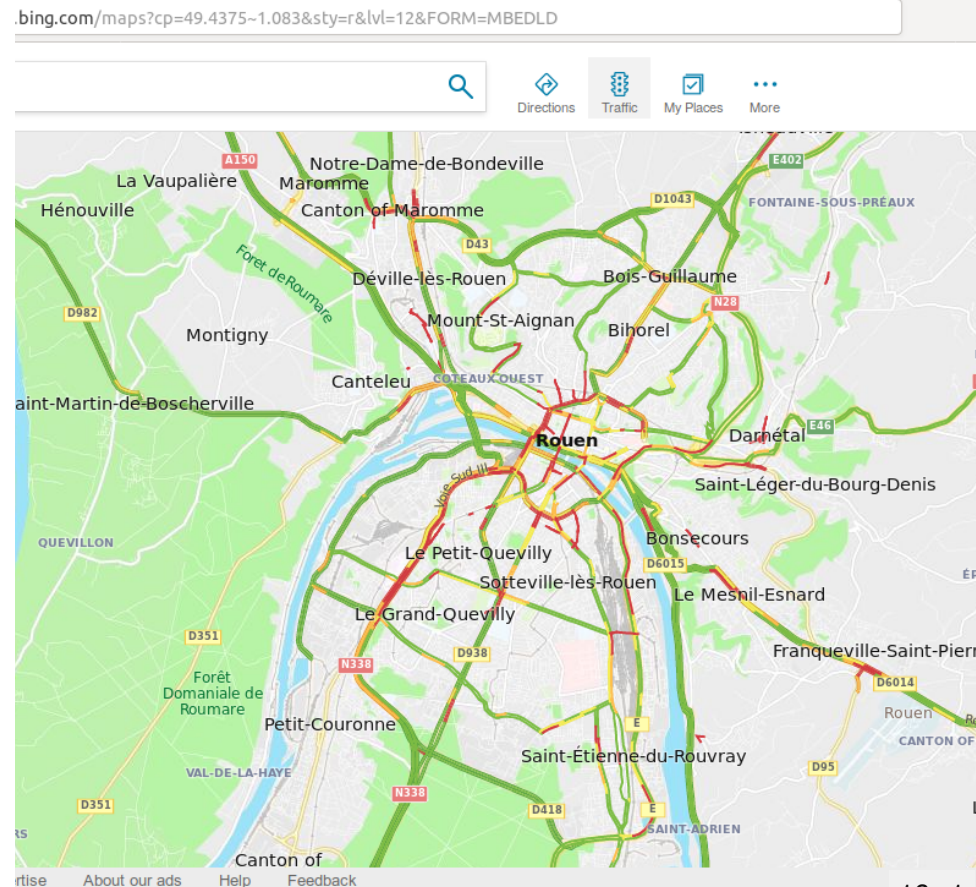


Condition de circulation



Bing Map (Microsoft)

Collecte en temps réel des conditions de circulation



Condition de circulation



Bing Map (Microsoft)

Collecte en temps réel des conditions de circulation

Site web : /www.bing.com/maps

Langage : R

Type de récolte : Campagne

Format : Raster (image)

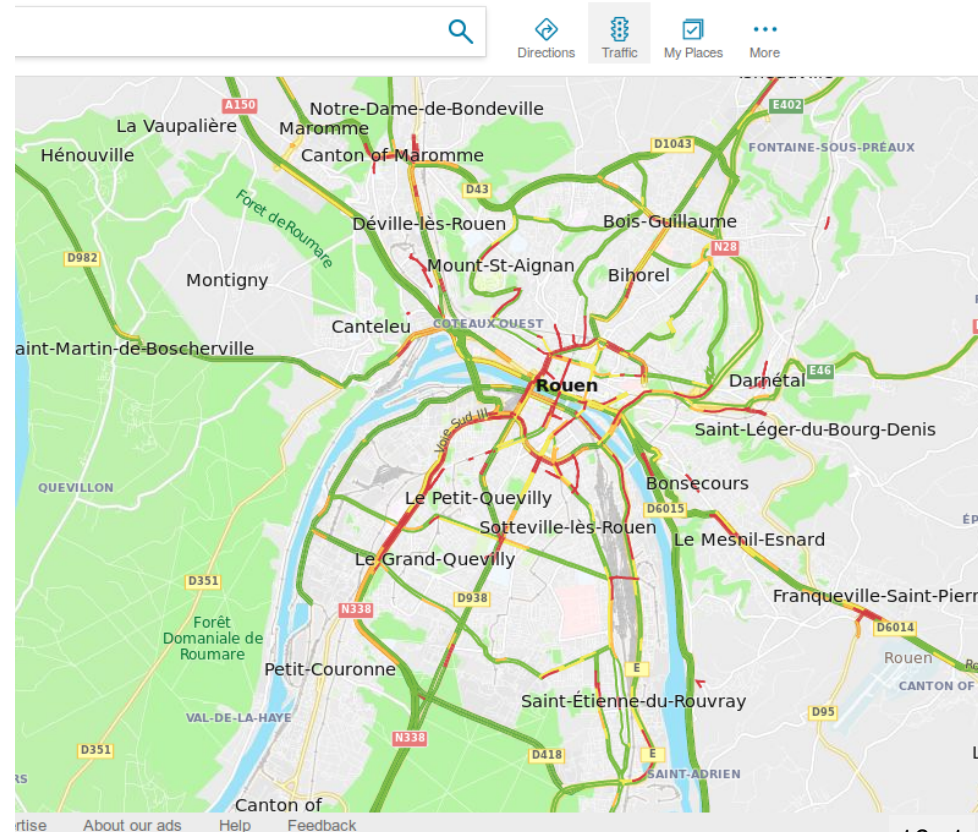
Avantages & limites :

- Trafic en temps réel
- Disponible dans 55 pays
- Méthodologie de Bing inconnue

Principe :

- Collecter et reprojeter des images
- Extraire les conditions de circulation

bing.com/maps?cp=49.4375~-1.083&sty=r&lvl=12&FORM=MBEDLD



Condition de circulation



Exemple : Bing Map (Microsoft)

Collecte en temps réel des conditions de circulation

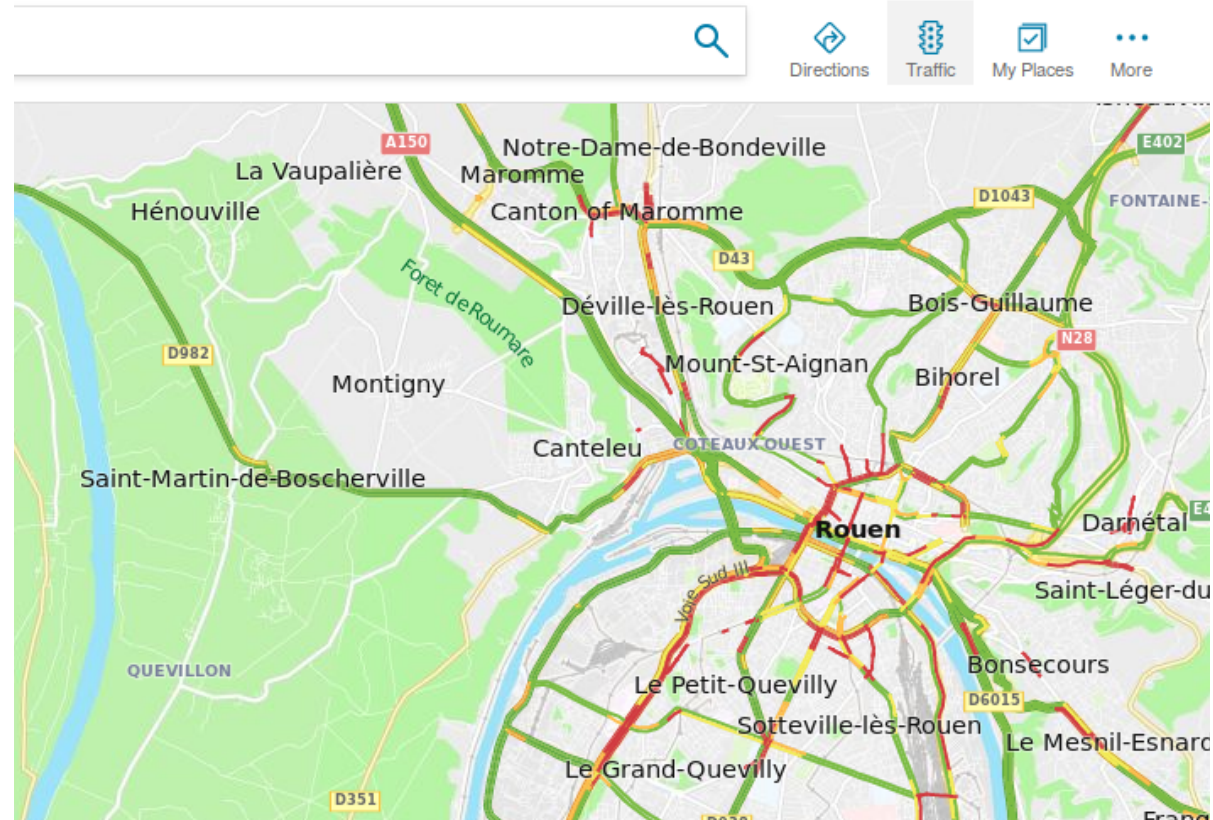
"As part of efforts on learning about traffic flows from data, researchers at Microsoft Research explored methods that enhance the safety and privacy of people who wish to help with the "crowdsourcing" of real-time flows of road data from their mobile GPS data. Principles of community sensing have been developed. These principles center on working with people under a "privacy budget" based on the use of the computations of the value of information for understanding flows over time on the road network"

<https://www.microsoft.com/en-us/research/project/predictive-analytics-for-traffic/?from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fprojects%2Fclearflow%2F>

Condition de trafic

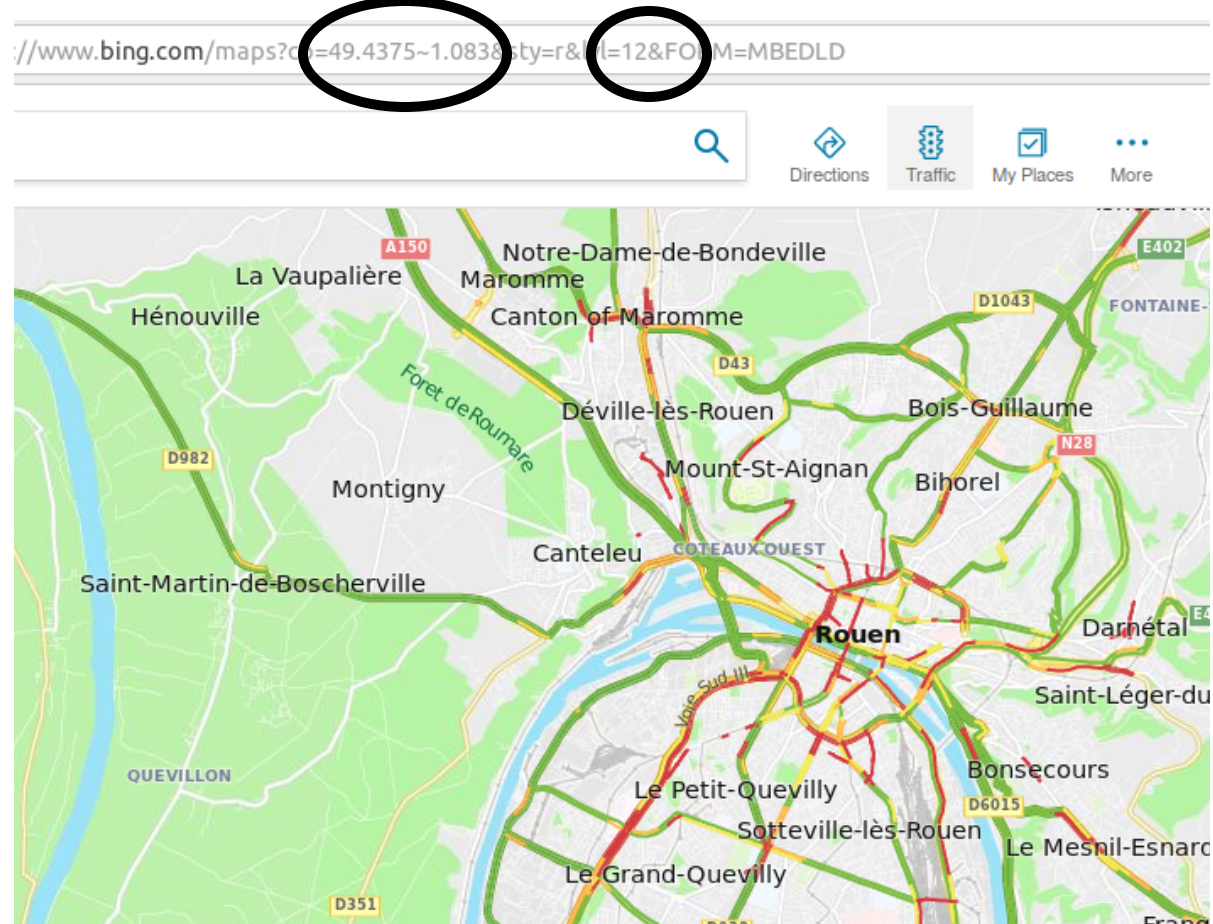
Première approche : capture d'écran

<https://www.bing.com/maps?cp=49.4375~-1.083&sty=r&lvl=12&FORM=MBEDLD>



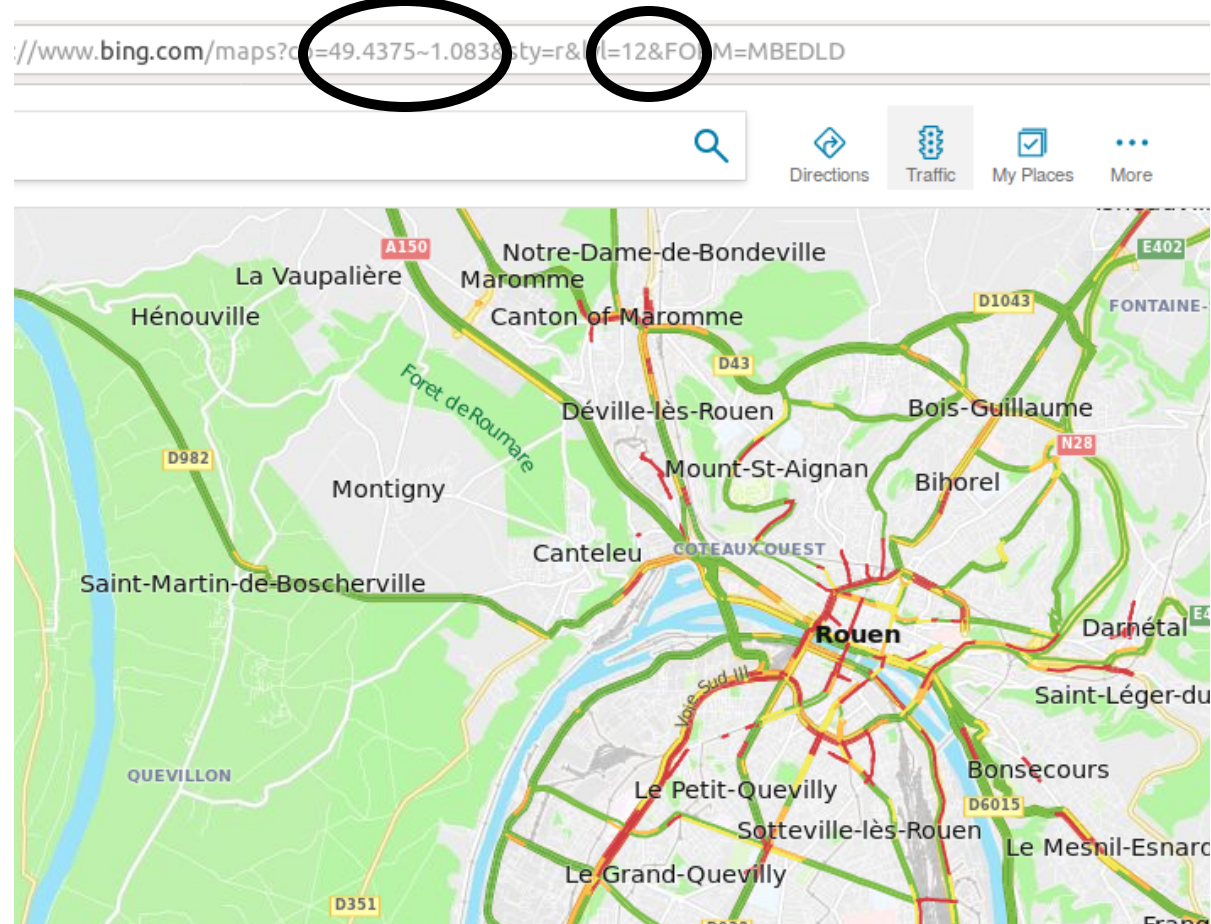
Condition de trafic

Première approche : capture d'écran



Condition de trafic

Première approche : capture d'écran



**Projeter l'image
connaissant :**

Centre de l'image :

Longitude = 1.083

Latitude = 49.4375

Niveau de zoom : 12

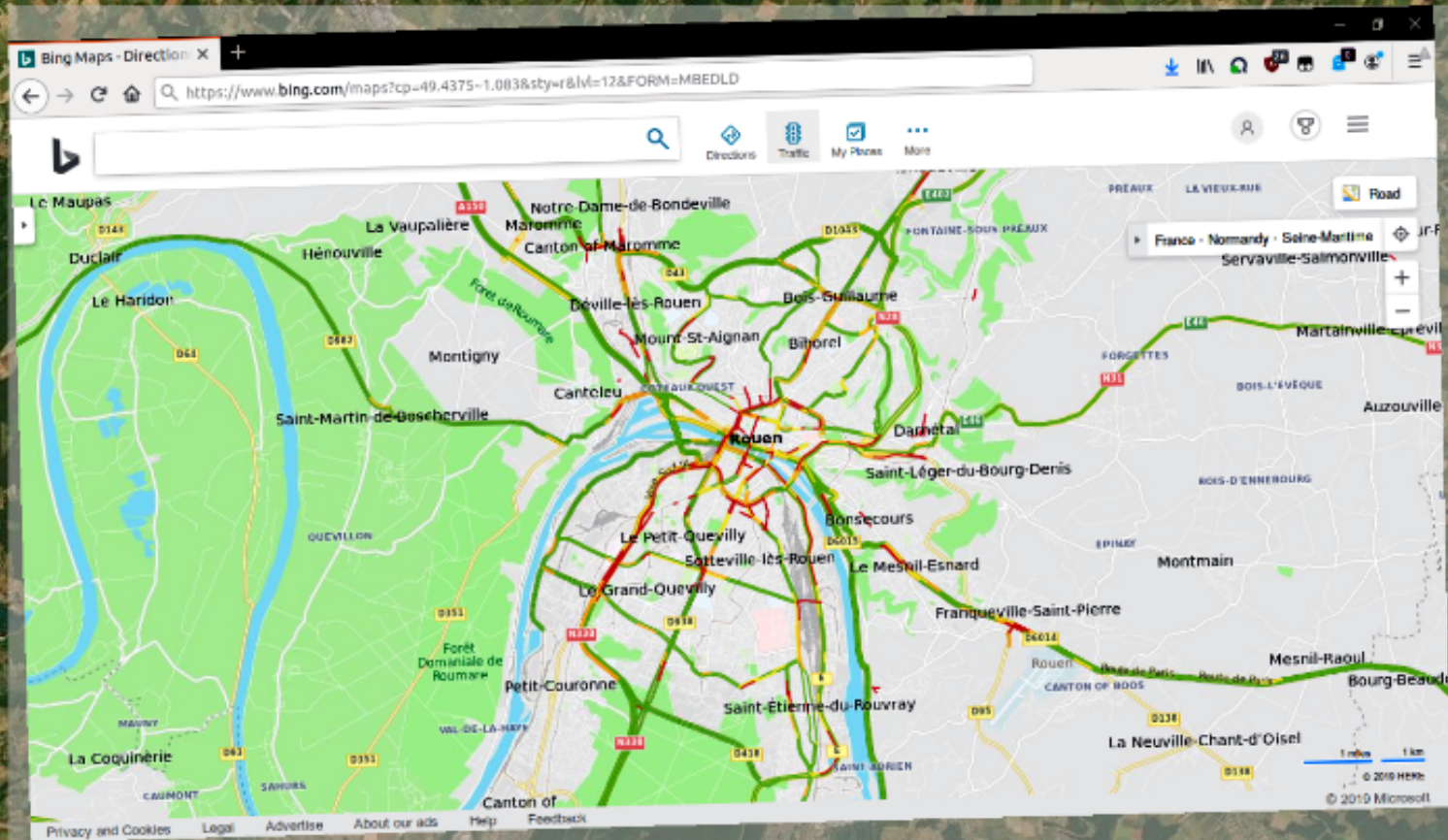
Projection Bing :

ici UTM 31N

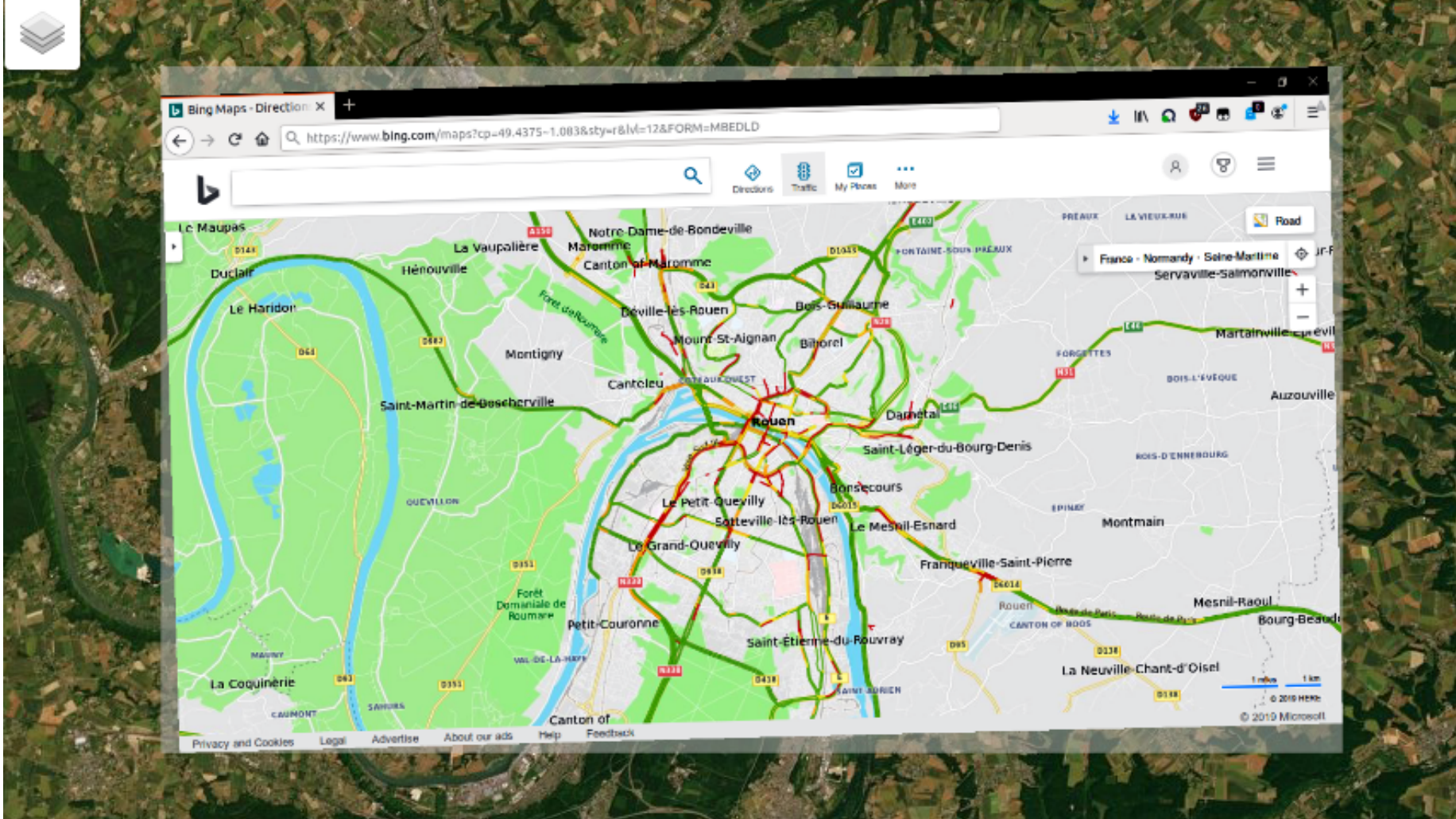
Résolution d'un pixel (m) :

$156543.04 * \cos(\text{lat}) / (2 \wedge \text{zoom})$

Condition de trafic



Condition de trafic



Prochaine étape : extraire les conditions de trafic

Condition de trafic

image avec trafic

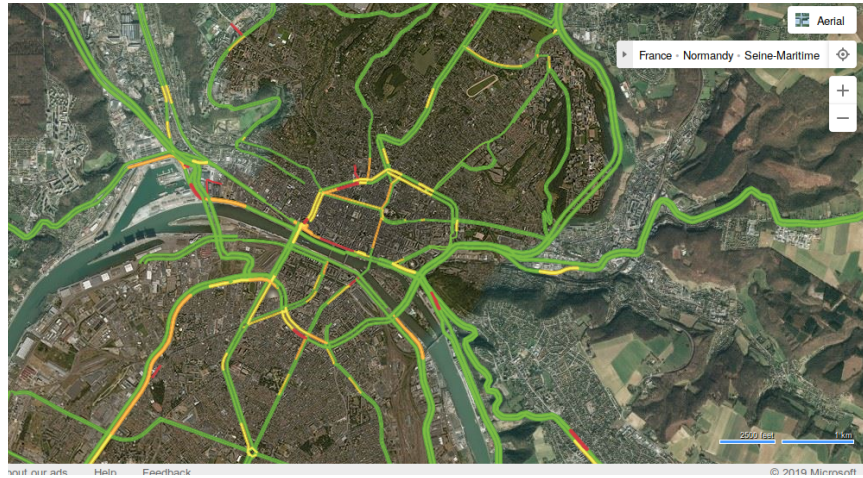
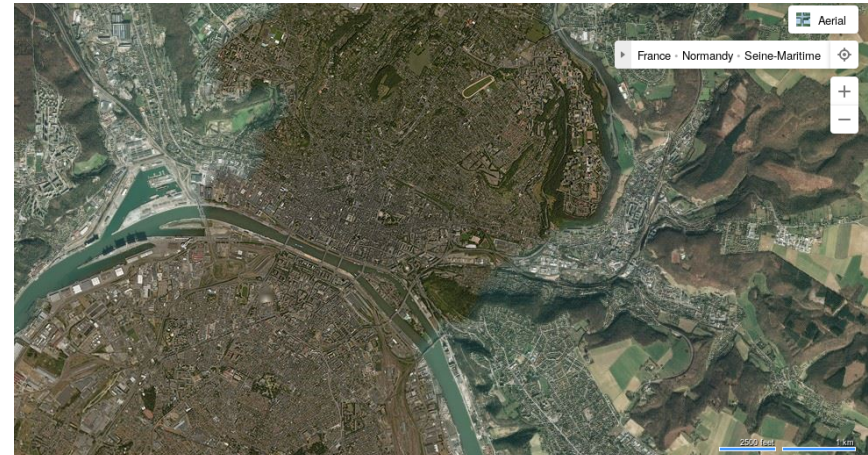


image avec sans trafic

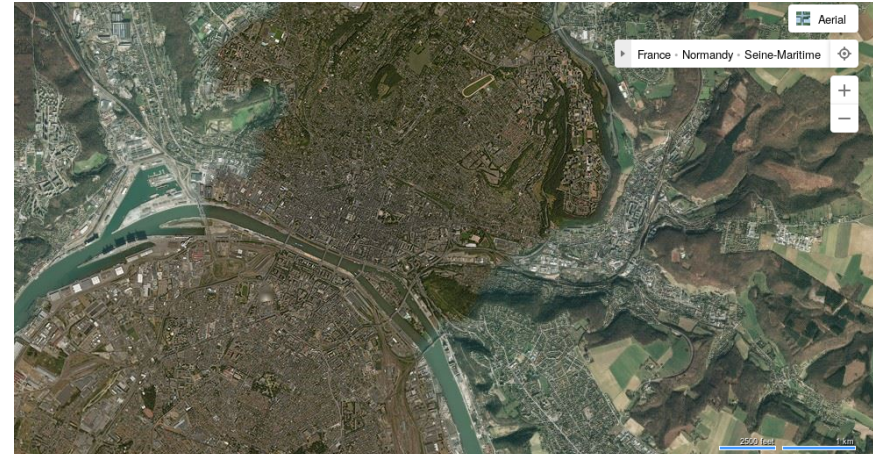


Condition de trafic

image avec trafic



image avec sans trafic

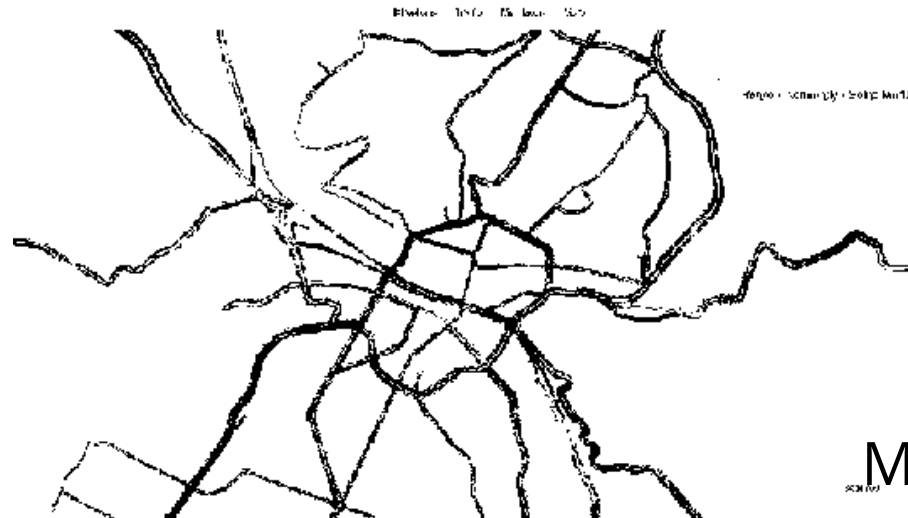


Condition de trafic

image avec trafic



image avec sans trafic



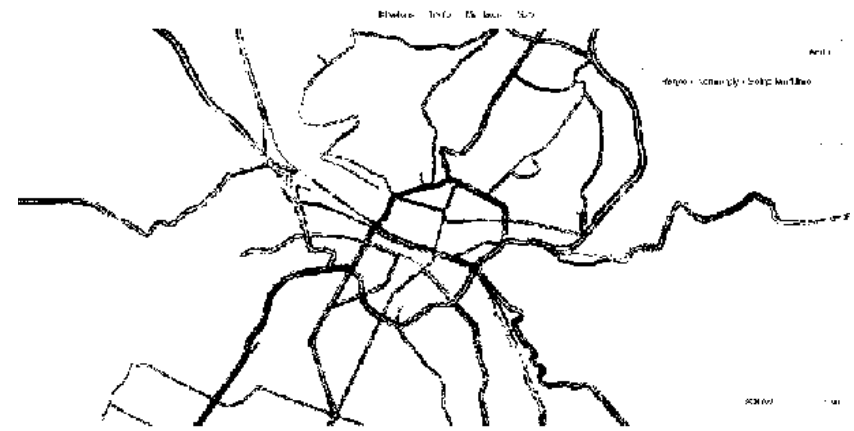
Masque

Condition de trafic

image avec trafic



Masque



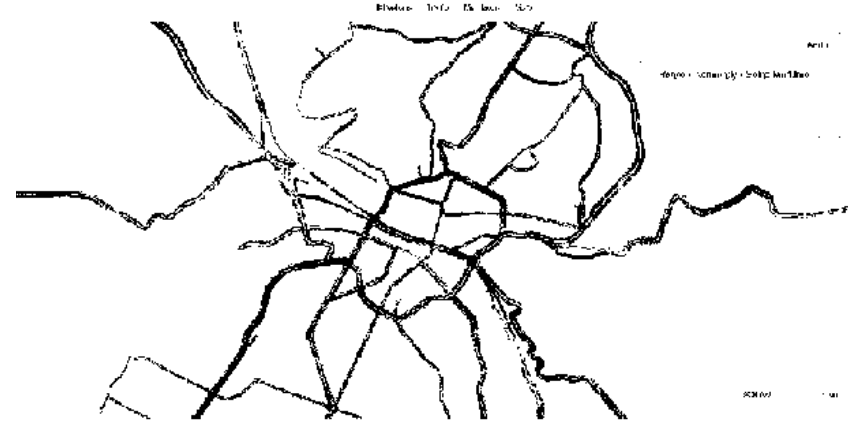
Condition de trafic

image avec trafic



Masque

X



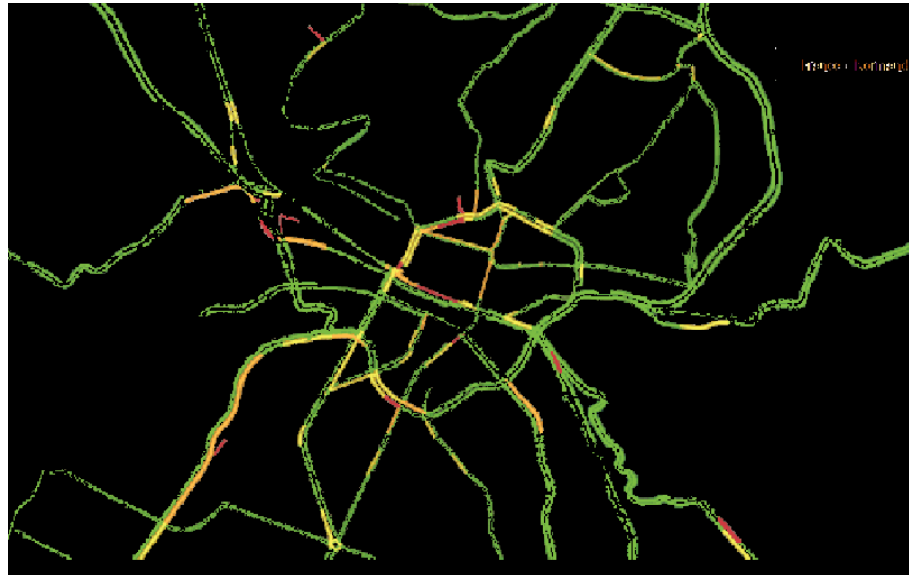
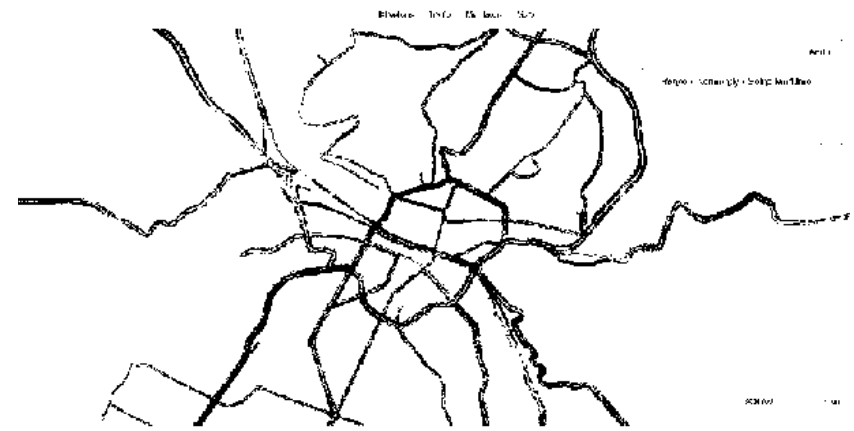
Condition de trafic

image avec trafic



Masque

X



(H) API !

Application Programming Interface

*“ [...] is a set of features and rules that exist inside a software program (the application) enabling interaction with it through software - as opposed to a human user interface. **The API can be seen as a simple contract (the interface) between the application offering it and other items, such as third party software or hardware.***

[src : developer.mozilla.org](http://developer.mozilla.org)



Condition de trafic

2ème approche : Utiliser l'API pour récupérer une carte statique

<https://dev.virtualsearth.net/REST/V1/Imagery/Mapstyle/lat,lon/zoom/?>

mapLayer=**TrafficFlow**&key=**Token**&autresparamètres

Adapter les paramètres pour obtenir une carte avec fond noir, sans texte :

Condition de trafic

2ème approche : Utiliser l'API pour récupérer une carte statique

<https://dev.virtualearth.net/REST/V1/Imagery/Mapstyle/lat,lon/zoom/?>

mapLayer=**TrafficFlow**&key=**Token**&autresparamètres

Adapter les paramètres pour obtenir une carte avec fond noir, sans texte :



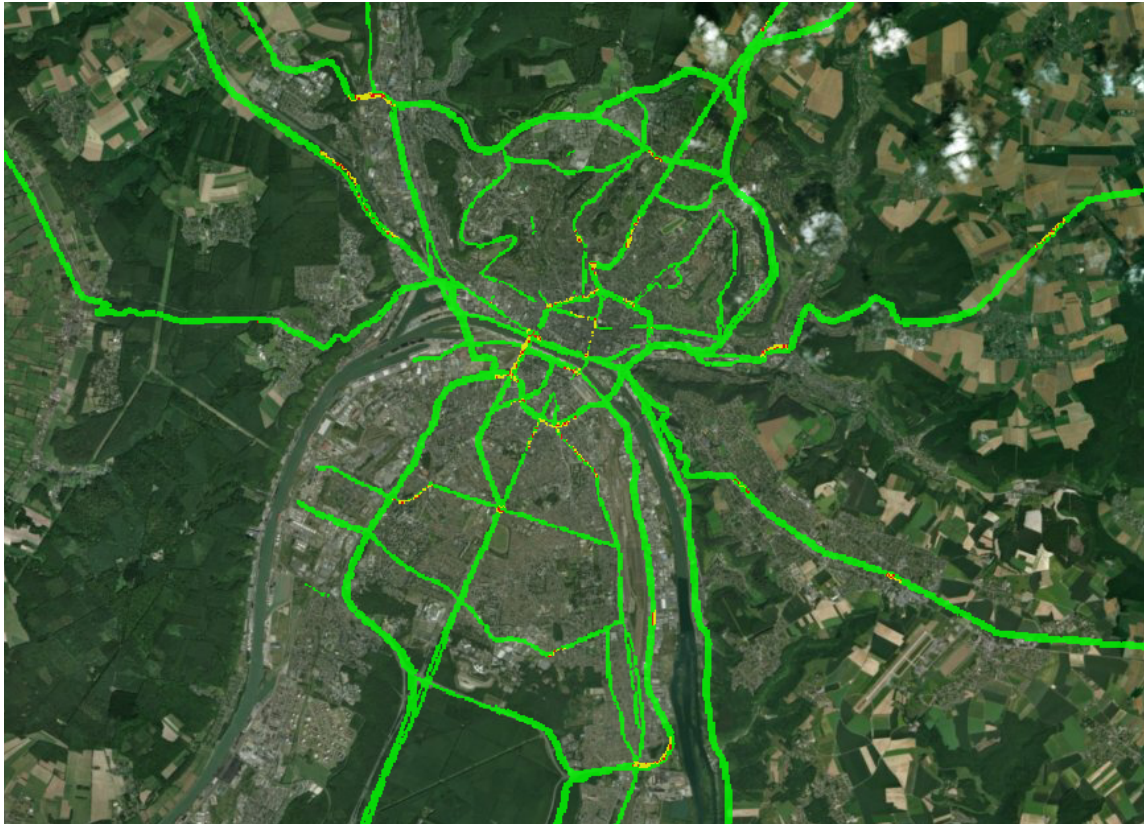
Condition de trafic

2ème approche : Utiliser l'API pour récupérer une carte statique

<https://dev.virtualearth.net/REST/V1/Imagery/Mapstyle/lat,lon/zoom/?>

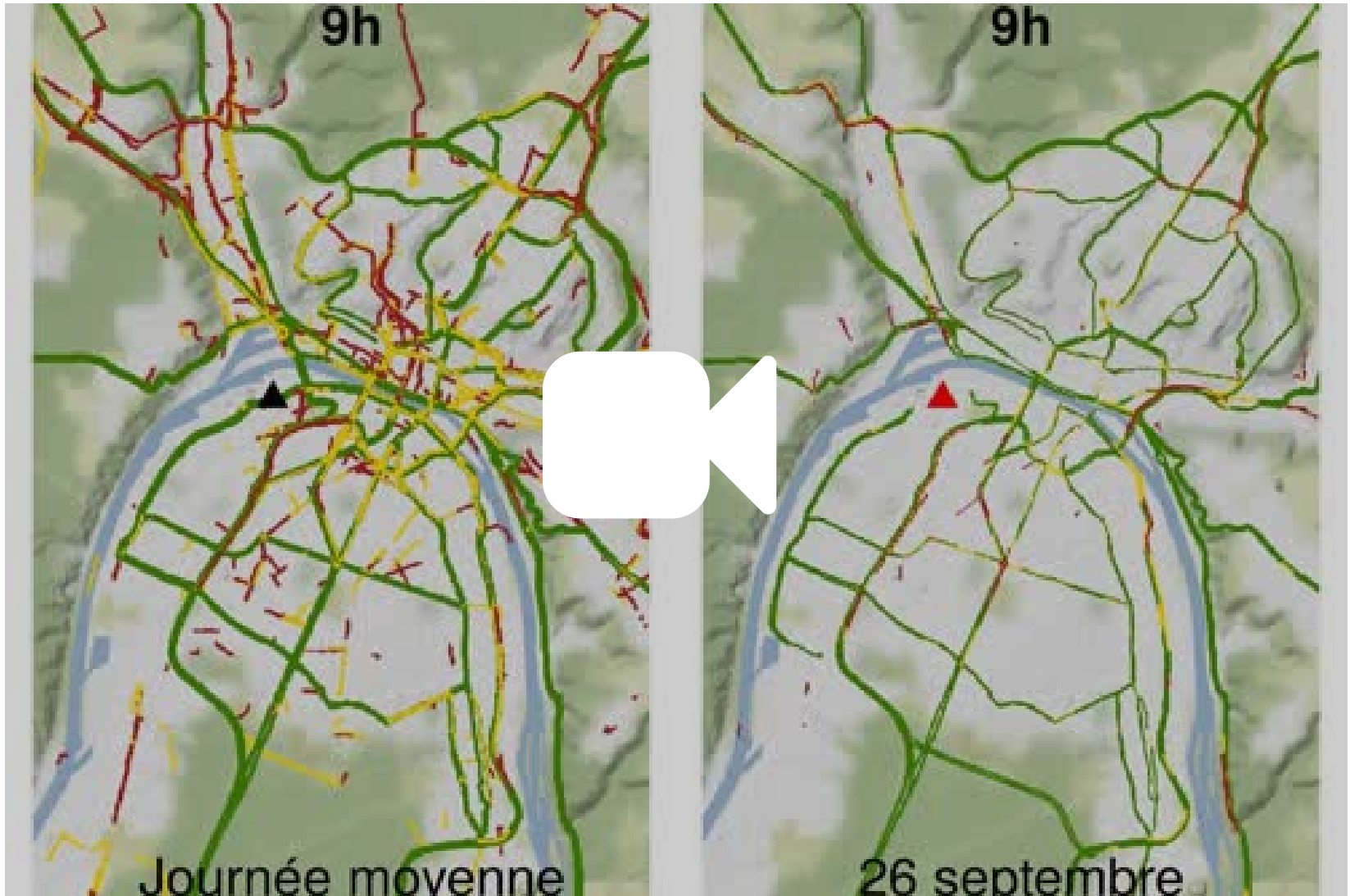
mapLayer=**TrafficFlow**&key=**Token**&autresparamètres

Adapter les paramètres pour obtenir une carte avec fond noir, sans texte :



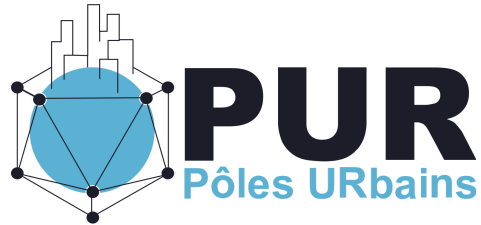
Sélection des pixels
vert, jaune et rouge

Condition de trafic



+ données comptages & enquêtes EMD

Arctique & système monde



ANR PUR (2016 - 2020) / Yvette Vaguet

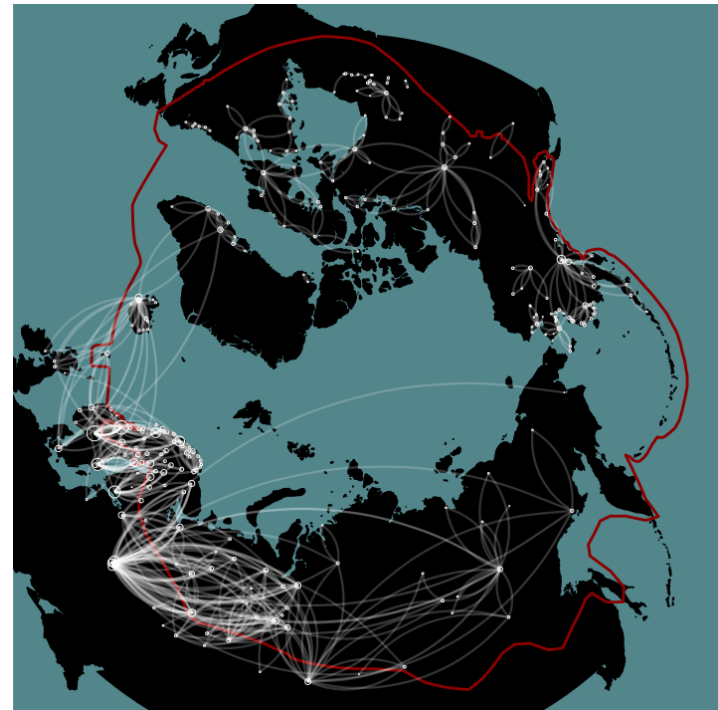
Intégration des zones urbaines arctiques dans le système monde

Un des objectifs :

Analyser temporellement les flux affectant les zones arctiques, notamment les flux aériens.

Comment ?

En créant des réseaux à partir des données les plus appropriées et accessibles



FlightRadar



flightradar 24

ANR PUR (2015 - 2019) / Yvette Vaguet

Récupérer les **départs et arrivées d'avions** pour les **aéroports mondiaux** pendant **plus d'un an**

Site web : flightradar24.com

Langage : Python

Framework : Scrapy

Type de récolte : Continue (depuis le **2018/06/10**)

Format : mongoDB / json

Difficultés :

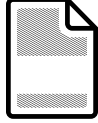
- architecture technique pour une récolte continue
- gestion des timezones
- intégration/consolidation des données SGBD (work in progress)



Free as in Freedom

<https://github.com/IDEES-Rouen/Flight-Scraping>

FlightRadar

API REST via appel XHR →  limité à 100 départs/arrivées
json

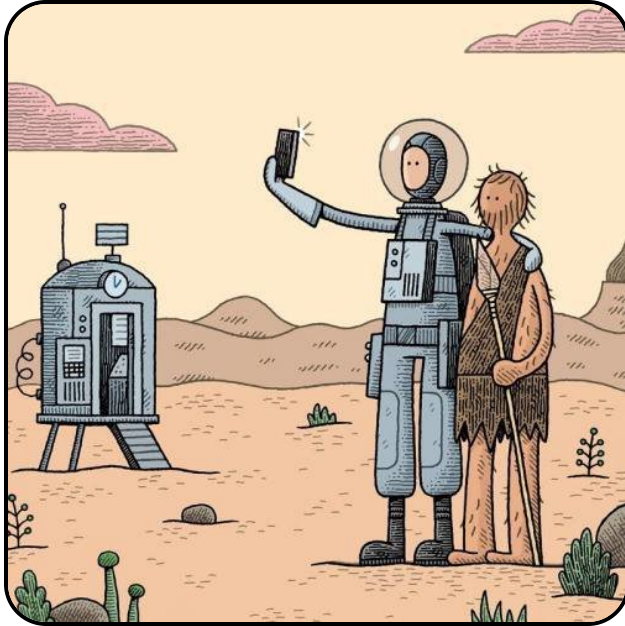
```
query =  
'https://api.flightradar24.com/common/v1/airport.json?  
code= {code} &plugin\[ \]=&plugin-setting\[schedule\[ \]  
[mode\[ \]=&plugin-setting\[schedule\[ \]\[timestamp\[ \]= {timestamp}  
&page= {page} &limit=100&token='
```

{code} = code aéroport IATA

{page} = numéro de la page

{timestamp} = date de la requête

h24 & 7j/7



src : tom gauld

Focus sur les données ...

Variabilité +/- grande des sources !

1 req. ~2 secondes

~ **4 heures** de collecte

2 jours en échecs

impacte

Structuration des données



json non
structuré

~ 20 Go / an
800Mo / jour

intégration : vérification,
consolidation




SGBD

+ bonus RPGD

h24 & 7j/7



YOU CAN TELL WHEN SOMEONE'S BEEN A PROGRAMMER FOR A WHILE BECAUSE THEY DEVELOP A DEEP-SEATED FEAR OF TIME ZONE PROBLEMS.

 zachholman.com

Time / TimeZone Nightmare !

Les jours font 24 h

... Heure d'été ...

Un jour fait 86400 secondes

... Secondes intercalaires ...

Les timezones sont stables

... NON ...

TZ Server = UTC

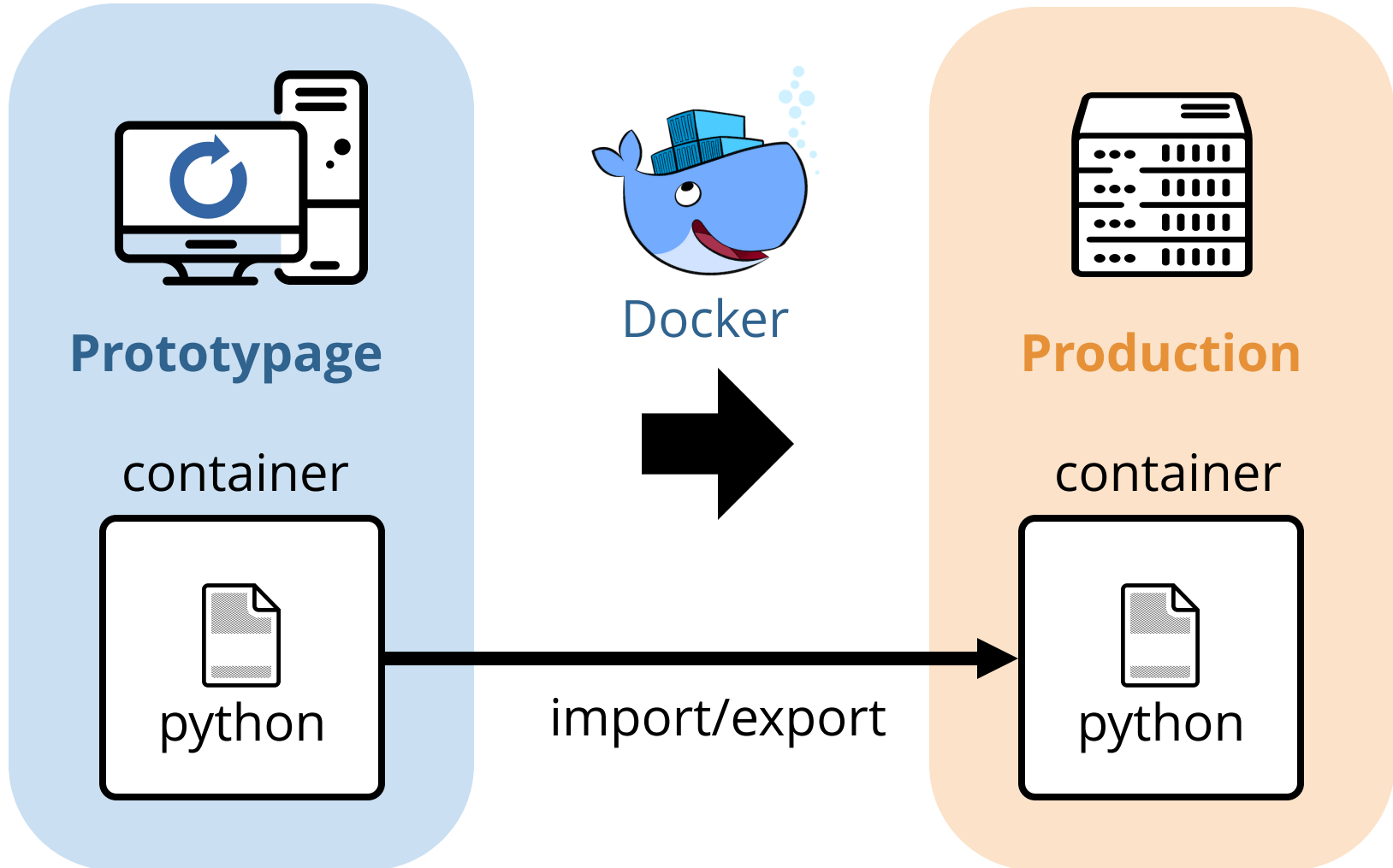
... Par convention oui mais ...

GMT = UTC

Non. UTC n'a pas de DST

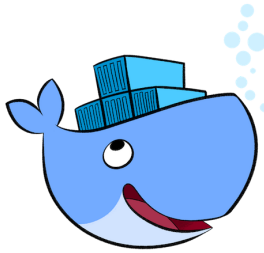
h24 & 7j/7

Quelle(s) infrastructure(s) pour y arriver ?



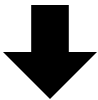
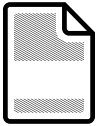
h24 & 7j/7

Quelle(s) infrastructure(s) pour y arriver ?

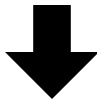
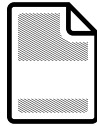


VM vs Docker ?

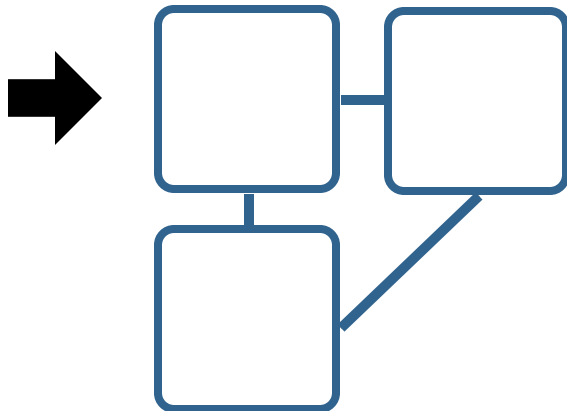
Dockerfile



docker-compose.yml



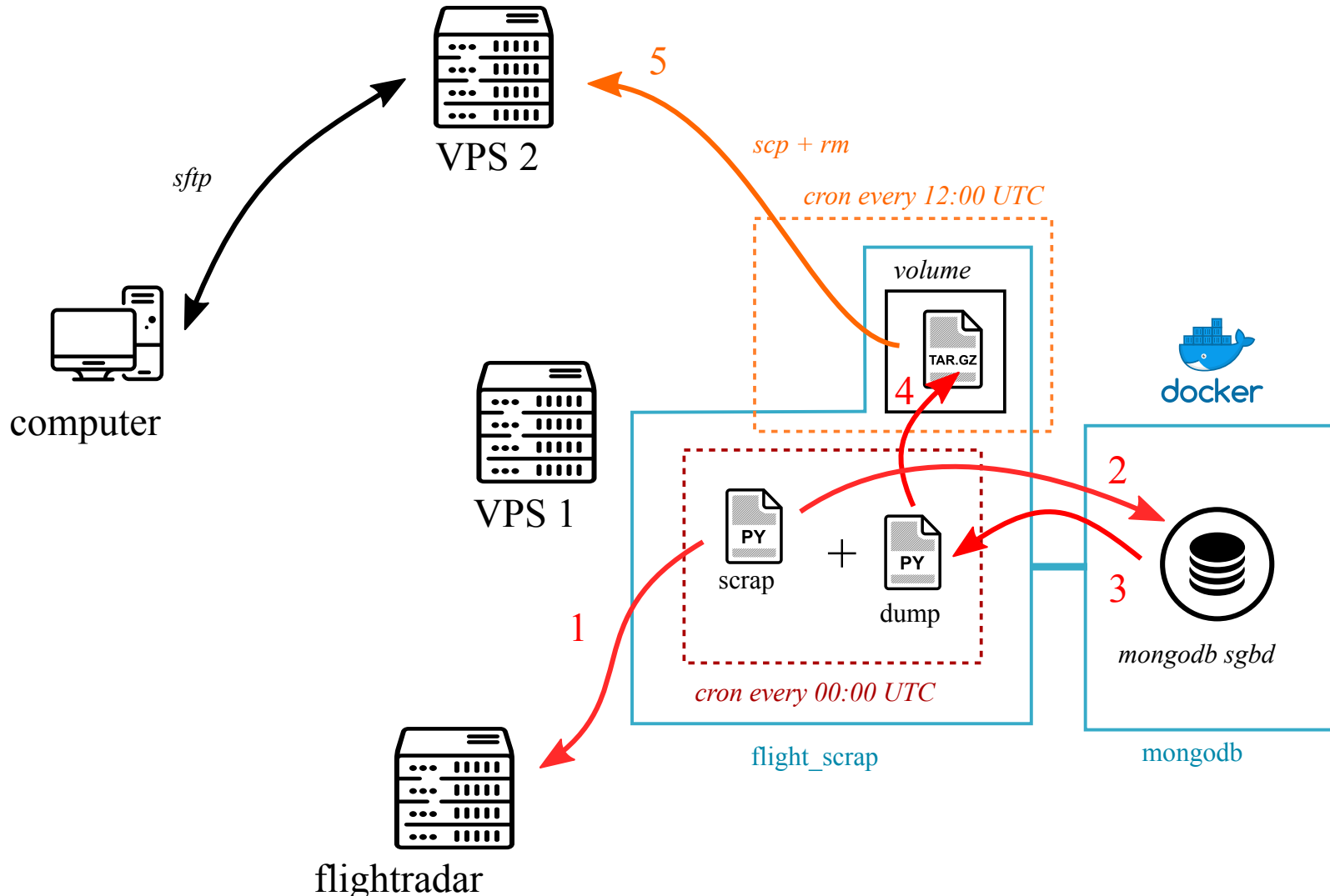
organisation logique



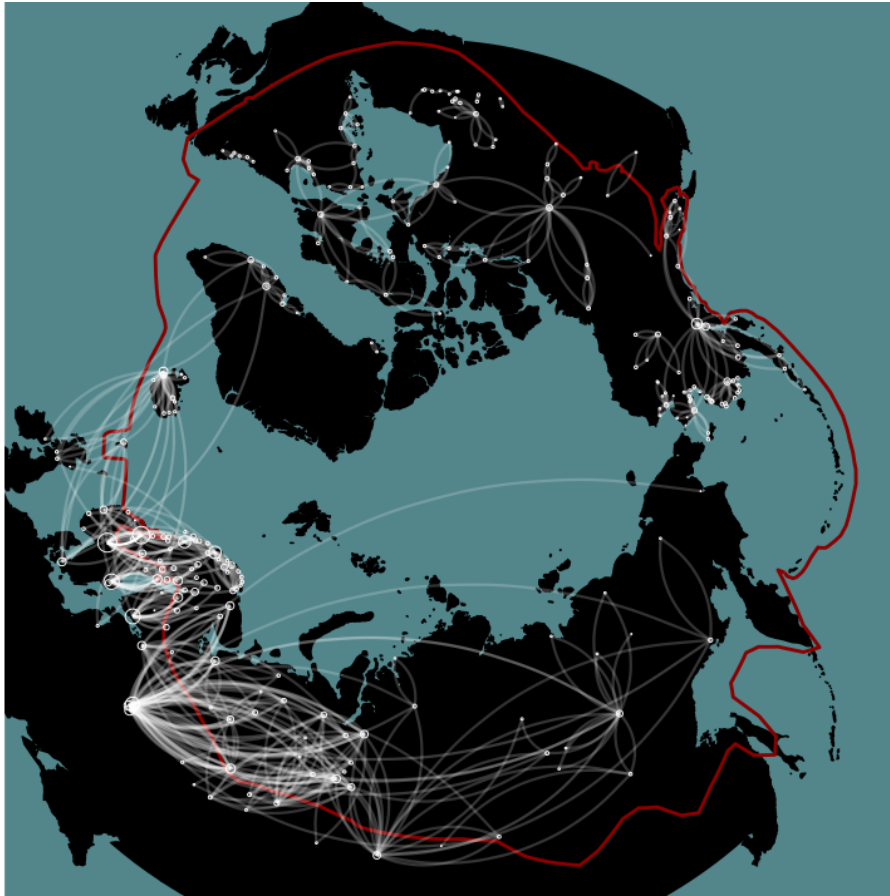
- micro-services (1 container = 1 application)
- plus simple
- plus flexible
- consomme moins de ressources
- facilite la **reproductibilité**

h24 & 7j/7

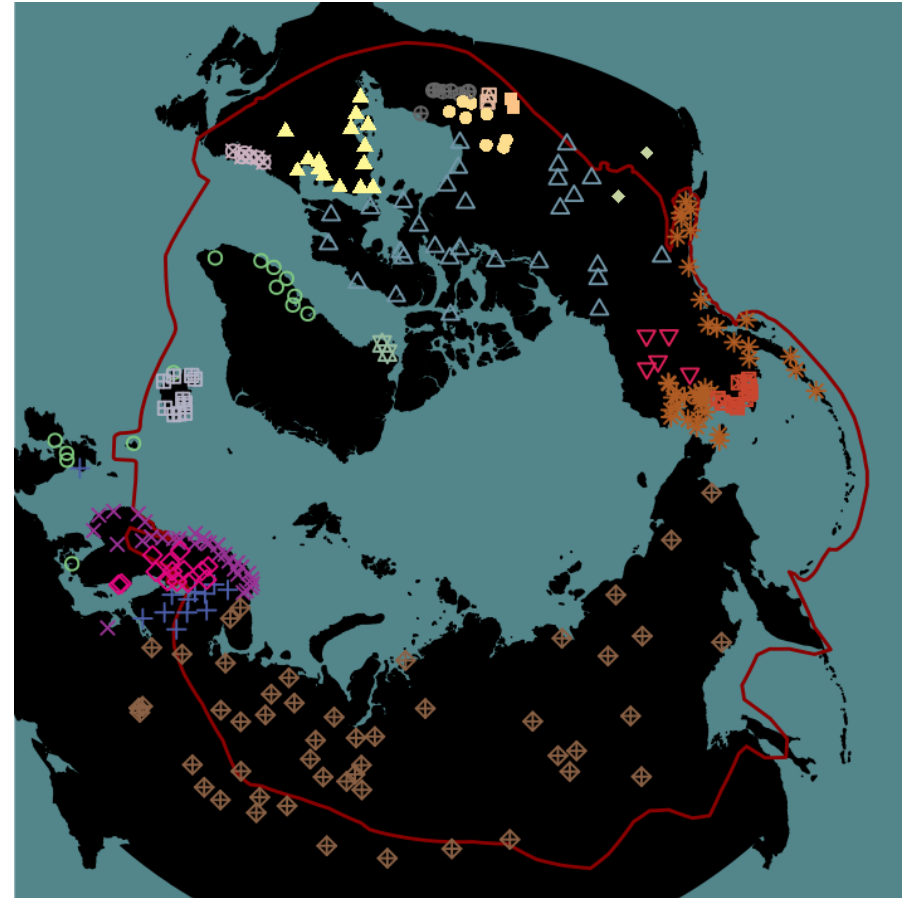
Quelle(s) infrastructure(s) pour y arriver ?



FlightRadar

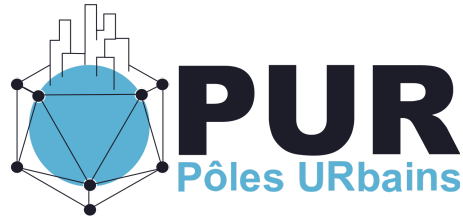


Vol vers la zone arctique



Clustering (louvain)

Arctique & système monde



ANR PUR (2016 - 2020) / Yvette Vaguet

Intégration des zones arctiques dans le système monde

Un des objectifs :

Analyser temporellement les flux affectant les zones arctiques, dont les flux touristiques.

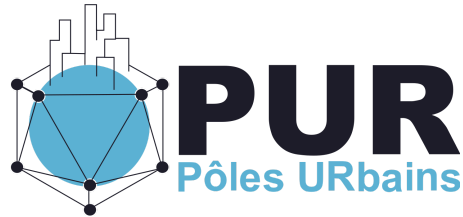
Pourquoi ?

"Tourism is a powerful global force that is turning the north into a Global North—tied into the Global South by facing the same global problem" (Veijola and Strauss-Mazzullo, 2019)

Comment ?

En créant des réseaux à partir des données les plus appropriées et accessibles

Arctique & système monde



ANR PUR (2016 - 2020) / Yvette Vaguet

Intégration des zones arctiques dans le système monde

Un des objectifs :

Analyser temporellement les flux affectant les zones arctiques, dont les flux touristiques.

Pourquoi ?

"Tourism is a powerful global force that is turning the north into a Global North—tied into the Global South by facing the same global problem" (Veijola and Strauss-Mazzullo, 2019)

Comment ?

En créant des réseaux à partir des données **7M+** les plus appropriées et accessibles

Airbnb listings worldwide

100K+

cities with Airbnb listings

191+

countries and regions with Airbnb listings



Type

Privately held company

Industry

Lodging

Founded

August 2008;
11 years ago in San Francisco, California

Founders

Brian Chesky
Joe Gebbia
Nathan Blecharczyk

Airbnb



Collecter des données permettant de former un réseau de flux touristiques vers l'Arctique

Site web : airbnb.com

Langage : R

Framework : rvest / httr

Type de récolte : Campagne

Format : json

Caractéristiques :

- Accessible via code source
- (+) Plusieurs API officielles & pas de token privé
- (-) 80 requêtes / min / IP ==> temps de collecte long

- Des extractions de zones (insideairbnb) mais que dans certaines villes
- Des codes sont disponibles (Tom Slee) python, mais pas sous R

Airbnb

Protocole :

1. Collecter les logements (listing) dans notre zone d'étude (fenêtres mobiles)

==> identifiant du propriétaire et du logement + localisation

2. Liste des personnes ayant visité un de ces logements

==> liste des personnes ayant laissé un commentaire

==> localisation du domicile (déclaré)

Airbnb

Protocole :

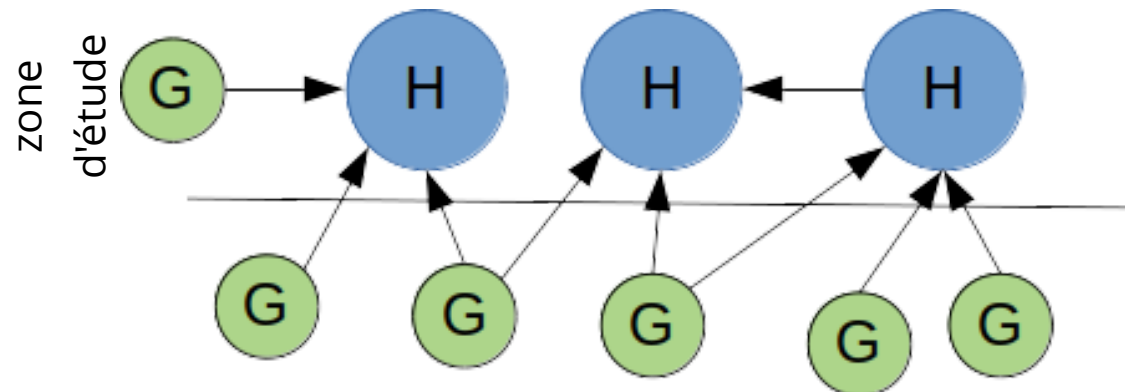
1. Collecter les logements (listing) dans notre zone d'étude (fenêtres mobiles)

==> identifiant du propriétaire et du logement + localisation

2. Liste des personnes ayant visité un de ces logements

==> liste des personnes ayant laissé un commentaire

==> localisation du domicile (déclaré)



H= Host (hôtebergeur)

G : Guest (visiteur)

Airbnb

Protocole :

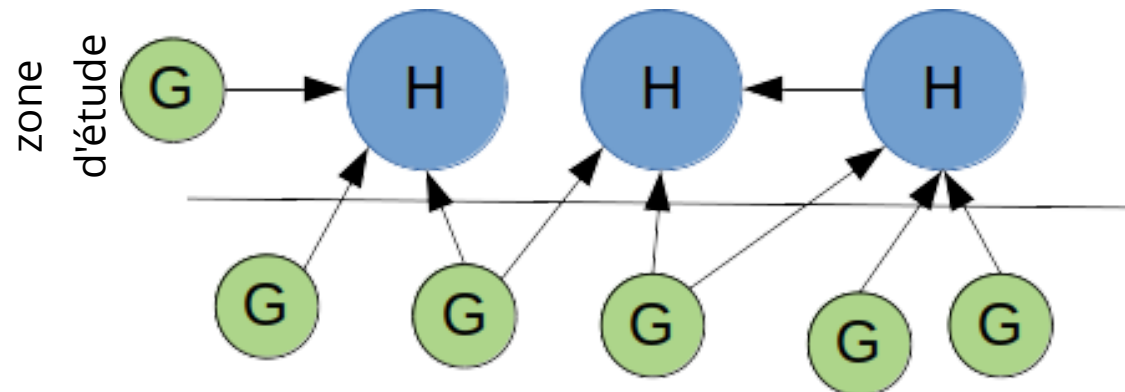
1. Collecter les logements (listing) dans notre zone d'étude (fenêtres mobiles)

==> identifiant du propriétaire et du logement + localisation

2. Liste des personnes ayant visité un de ces logements

==> liste des personnes ayant laissé un commentaire

==> localisation du domicile (déclaré)



3. Collecter tous les commentaires concernant ces personnes.

==> Liste des personnes accueillies ou hébergées (+ localisation)

H= Host (hôte/bergeur)

G : Guest (visiteur)

Airbnb

Protocole :

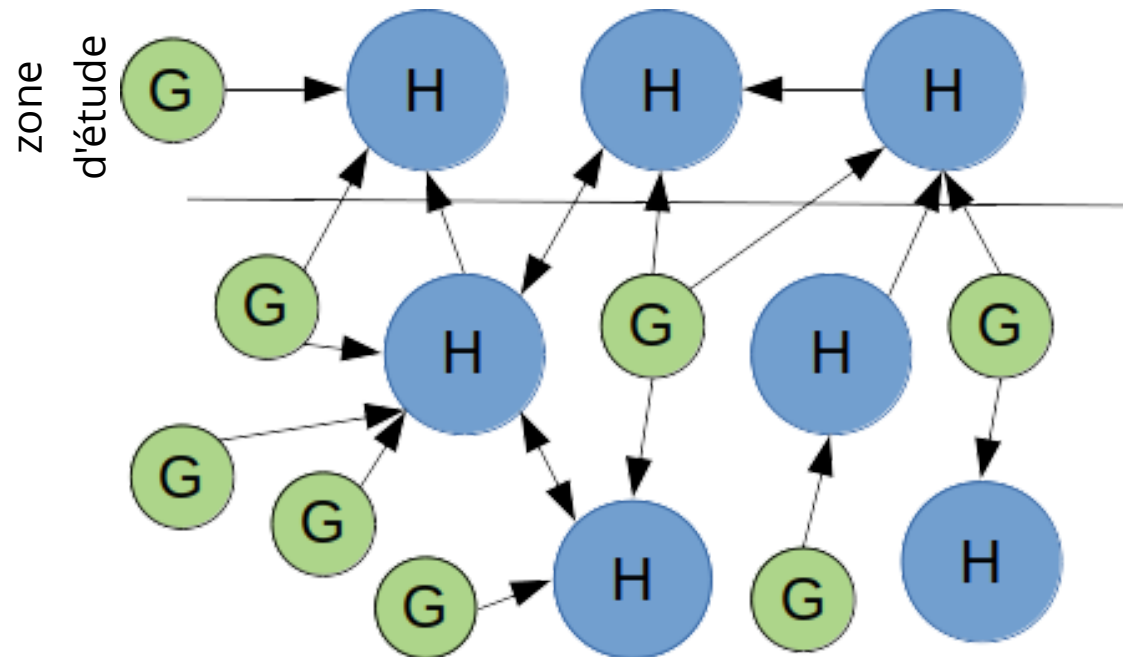
1. Collecter les logements (listing) dans notre zone d'étude (fenêtres mobiles)

==> identifiant du propriétaire et du logement + localisation

2. Liste des personnes ayant visité un de ces logements

==> liste des personnes ayant laissé un commentaire

==> localisation du domicile (déclaré)



3. Collecter tous les commentaires concernant ces personnes.

==> Liste des personnes accueillies ou hébergées (+ localisation)

H= Host (hôte/bergeur)

G : Guest (visiteur)

Airbnb : Listing

Dates Guests Work trip Type of place Price Instant Book More filters

8 places to stay

SUPERHOST Private room

★ 4.87 (209)

Geysir Hestar private room

2 guests · 1 bedroom · 2 beds · 4 baths
Free parking · Wifi



SUPERHOST Entire bungalow

★ 4.99 (191)

Horse Breeding Farm Jaðar

5 guests · 2 bedrooms · 3 beds · 1 bath
Kitchen · Free parking · Wifi



Show Map

Search as I move the map



Liste de
logements dans
une fenêtre
géographique
donnée

Terms, Privacy, Currency & More

Map data ©2019 Terms of Use Report a map error

Airbnb : Listing

Via l'API :

Dates Guests Work trip Type of place Price Instant Book More filters

8 places to stay

SUPERHOST Private room

★ 4.87 (209)

Geysir Hestar private room

2 guests · 1 bedroom · 2 beds · 4 baths
Free parking · Wifi



SUPERHOST Entire bungalow

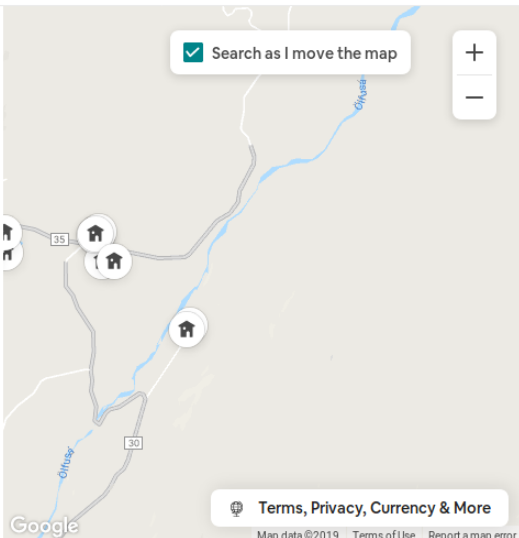
★ 4.99 (191)

Horse Breeding Farm Jaðar

5 guests · 2 bedrooms · 3 beds · 1 bath
Kitchen · Free parking · Wifi



Show Map



Liste de logements dans une fenêtre géographique donnée

listing:

```
badges: []
bathroom_label: "1 bath"
bathrooms: 1
bed_label: "3 beds"
bedroom_label: "2 bedrooms"
bedrooms: 2
beds: 3
city: ""
guest_label: "5 guests"
host_languages: [...]
id: 8513301
is_business_travel_ready: false
is_new_listing: false
is_superhost: true
kicker_content: {}
lat: 64.28767
lng: -20.16458
localized_city: "IS"
name: "Horse Breeding Farm Jaðar"
person_capacity: 5
picture_count: 41
```

Airbnb : Listing

Via l'API :

Dates Guests Work trip Type of place Price Instant Book More filters

8 places to stay

SUPERHOST Private room

★ 4.87 (209)

Geysir Hestar private room

2 guests · 1 bedroom · 2 beds · 4 baths
Free parking · Wifi



SUPERHOST Entire bungalow

★ 4.99 (191)

Horse Breeding Farm Jaðar

5 guests · 2 bedrooms · 3 beds · 1 bath
Kitchen · Free parking · Wifi



Show Map

Search as I move the map



Liste de
logements dans
une fenêtre
géographique
donnée

▼ listing:

```
badges: []
bathroom_label: "1 bath"
bathrooms: 1
bed_label: "3 beds"
bedroom_label: "2 bedrooms"
bedrooms: 2
beds: 3
city: ""
guest_label: "5 guests"
host_languages: [...]
id: 8513301
is_business_travel_ready: false
is_new_listing: false
is_superhost: true
kicker_content: {}
lat: 64.28767
lng: -20.16458
localized_city: "IS"
name: "Horse Breeding Farm Jaðar"
person_capacity: 5
picture_count: 41
```

Airbnb : Listing

Via l'API :

Dates Guests Work trip Type of place Price Instant Book More filters

8 places to stay

SUPERHOST Private room

★ 4.87 (209)

Geysir Hestar private room

2 guests · 1 bedroom · 2 beds · 4 baths
Free parking · Wifi



SUPERHOST Entire bungalow

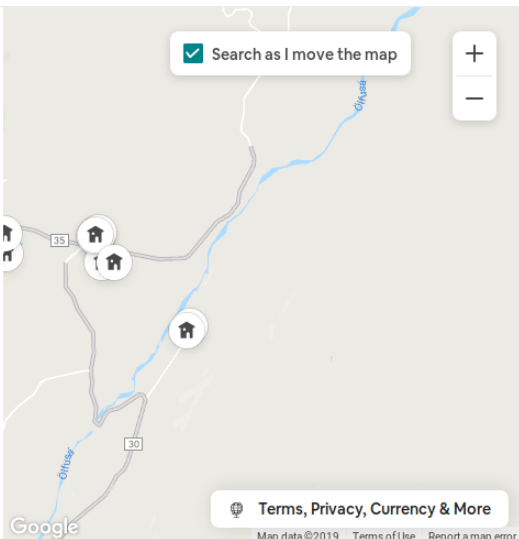
★ 4.99 (191)

Horse Breeding Farm Jaðar

5 guests · 2 bedrooms · 3 beds · 1 bath
Kitchen · Free parking · Wifi



Show Map



Liste de logements dans une fenêtre géographique donnée

listing:

```
badges: []
bathroom_label: "1 bath"
bathrooms: 1
bed_label: "3 beds"
bedroom_label: "2 bedrooms"
bedrooms: 2
beds: 3
city: ""
guest_label: "5 guests"
host_languages: [...]
id: 8513301
is_business_travel_ready: false
is_new_listing: false
is_superhost: true
kicker_content: {...}
lat: 64.28767
lng: -20.16458
localized_city: "IS"
name: "Horse Breeding Farm Jaðar"
person_capacity: 5
picture_count: 41
```

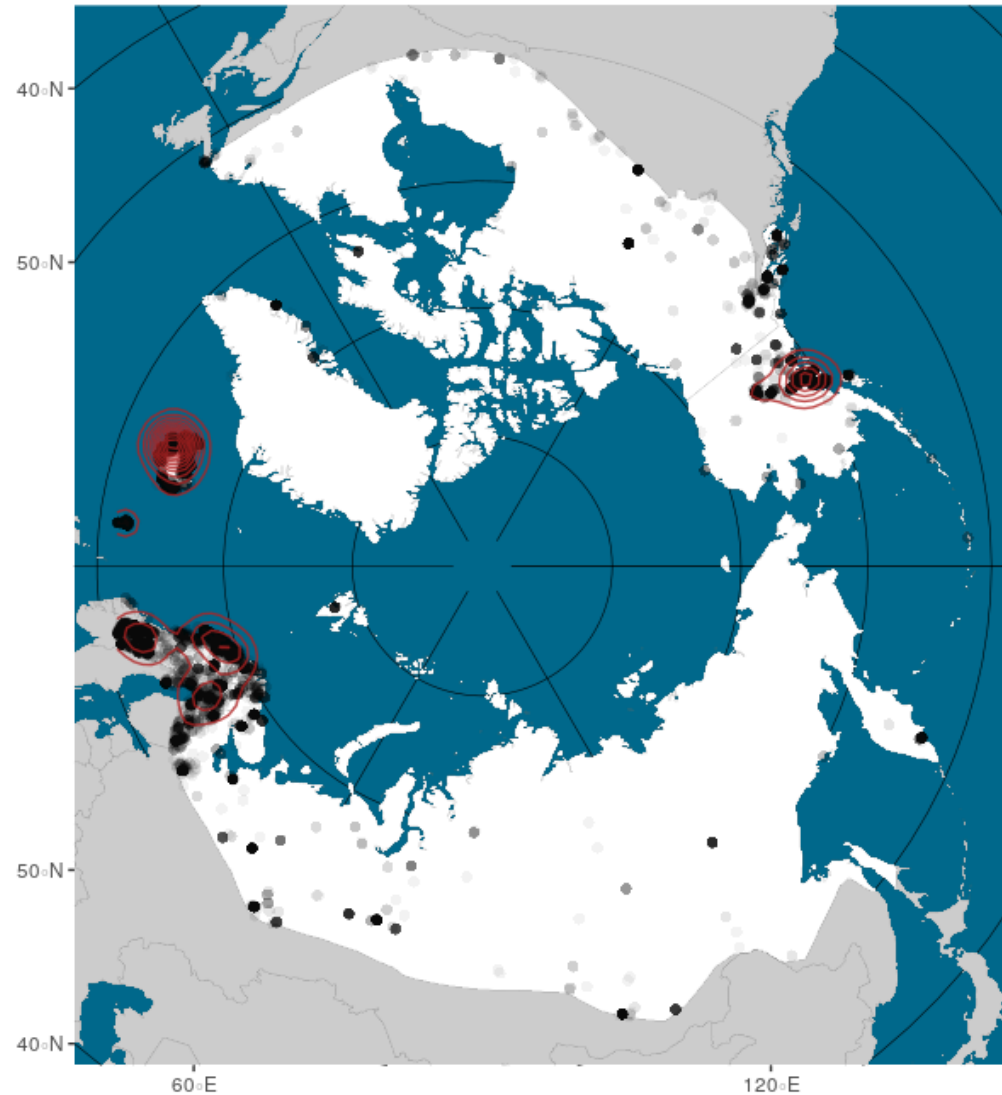
=> Collecte des informations
ex : host id, place id, lon/lat, etc.

Airbnb

Première étape :

Collecter les logements (listing)
dans notre zone d'étude.
(fenêtres mobiles)

- $n=24K$
- durée ~ 2 jours



Airbnb : commentaires

Page du logement

 <https://www.airbnb.com/rooms/8513301> Log in



Horse Breeding Farm Jaðar

IS



Kristbjörg

5 guests 2 bedrooms 3 beds 1 bath

Entire home
You'll have the bungalow to yourself.

Add dates for prices

4.99 (191 reviews)

Add dates for exact pricing


Dates


Check-in → Checkout

Guests


Airbnb : commentaires

Page du logement

 <https://www.airbnb.com/rooms/8513301> Log in



Horse Breeding Farm Jaðar
IS

 Kristbjörg

5 guests 2 bedrooms 3 beds 1 bath

Entire home
You'll have the bungalow to yourself.

Add dates for prices
★4.99 (131 reviews)

Add dates for exact pricing

Dates

Check-in → Checkout

Guests

Commentaires :

Stayed at Horse Breeding Farm Jaðar

August 2016



Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA

Joined in 2008 ·

Airbnb : commentaires

Page du logement

<https://www.airbnb.com/rooms/8513301> p Log in



Horse Breeding Farm Jaðar

IS



Kristbjörg

5 guests 2 bedrooms 3 beds 1 bath

Entire home
You'll have the bungalow to yourself.

Add dates for prices
★4.99 (191 reviews)

Add dates for exact pricing

Dates

Check-in → Checkout

Guests

Commentaires :

Stayed at Horse Breeding Farm Jaðar

August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA

Joined in 2008 ·


Via l'API :


```
▼ comments: "Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend! "
```

```
  created_at: "2016-08-31T10:19:19Z"
  hide_author_profile: false
  id: 98520278
  language: "en"
  ▶ listing: {...}
  ▶ localized_review: {...}
  ▶ reviewee: {...}
  ▼ reviewer:
    created_at: "2008-03-02T06:20:00Z"
    deleted: false
    id: 2
    is_superhost: false
    first_name: "Joe"
    last_name: "prune_me"
    location: "San Francisco, CA"
    ▶ picture_url: "https://a0.muscache.com/..._policy=profile_x_medium"
    profile_path: "/users/show/2"
    ▶ profile_pic_path: "https://a0.muscache.com/..._policy=profile_x_medium"
    response: ""
    role: "guest"
```


Airbnb : commentaires

Page du logement


 <https://www.airbnb.com/rooms/8513301> Log in



Horse Breeding Farm Jaðar
IS

5 guests 2 bedrooms 3 beds 1 bath

Entire home
You'll have the bungalow to yourself.

 Kristbjörg

Add dates for prices
★4.99 (131 reviews)

Add dates for exact pricing

Dates
Check-in → Checkout

Guests

View Photos

Commentaires :

Stayed at Horse Breeding Farm Jaðar

August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008 ·

Via l'API :

```
▼ comments: "Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend! "
```


```
  created_at: "2016-08-31T10:19:19Z"
  hide_author_profile: false
  id: 98520278
  language: "en"
  ▶ listing: {...}
  ▶ localized_review: {...}
  ▶ reviewee: {...}
  ▼ reviewer:
    created_at: "2008-03-02T06:20:00Z"
    deleted: false
    id: 2
    is_superuser: false
    first_name: "Joe"
    last_name: "prune_me"
    location: "San Francisco, CA"
  ▶ picture_url: "https://a0.muscache.com/_policy=profile_x_medium"
  profile_path: "/users/show/2"
  ▶ profile_pic_path: "https://a0.muscache.com/_policy=profile_x_medium"
  response: ""
  role: "guest"
```

==> Date, identifiant, rôle, lieu déclaré, de toutes les personnes ayant commenté

Airbnb : commentaires

Page du logement

<https://www.airbnb.com/rooms/8513301> Log in



Horse Breeding Farm Jaðar
IS

5 guests 2 bedrooms 3 beds 1 bath

Entire home
You'll have the bungalow to yourself.

Kristbjörg

Add dates for prices
4.99 (131 reviews)

Add dates for exact pricing

Dates
Check-in → Checkout

Guests

View Photos

Commentaires :

Stayed at Horse Breeding Farm Jaðar

August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008

==> Date, identifiant, rôle, lieu déclaré, de toutes les personnes ayant commenté

Via l'API :

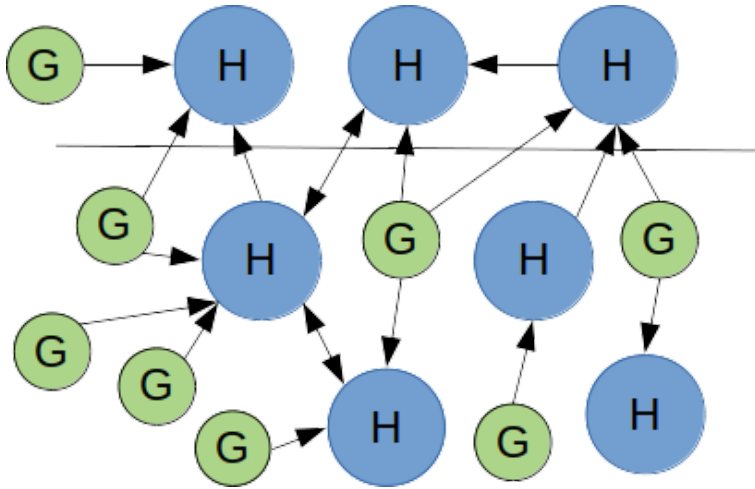
```
▼ comments: "Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend! "
```

```
created_at: "2016-08-31T10:19:19Z"  
hide_author_profile: false  
id: 98520278  
language: "en"  
▶ listing: {...}  
▶ localized_review: {...}  
▶ reviewee: {...}  
▼ reviewer:  
  created_at: "2008-03-02T06:20:00Z"  
  deleted: false  
  id: 2  
  is_superuser: false  
  first_name: "Joe"  
  last_name: "prune_me"  
  location: "San Francisco, CA"  
  ▶ picture_url: "https://a0.muscache.com/_policy=profile_x_medium"  
  profile_path: "/users/show/2"  
  ▶ profile_pic_path: "https://a0.muscache.com/_policy=profile_x_medium"  
  response: ""  
  role: "guest"
```

Idem pour la page de l'utilisateur : informations sur les visiteurs ou visités

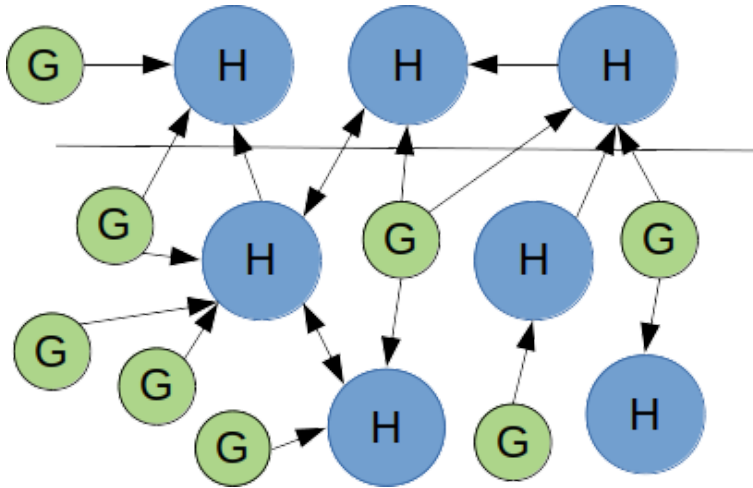
Airbnb

Création de réseau :



Airbnb

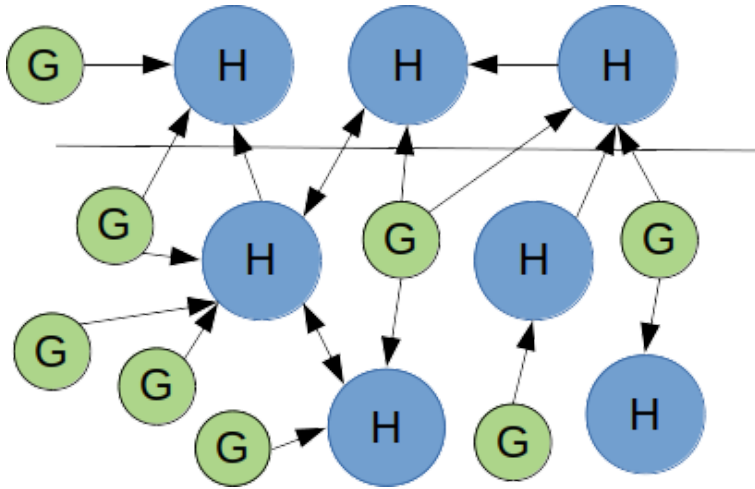
Création de réseau :



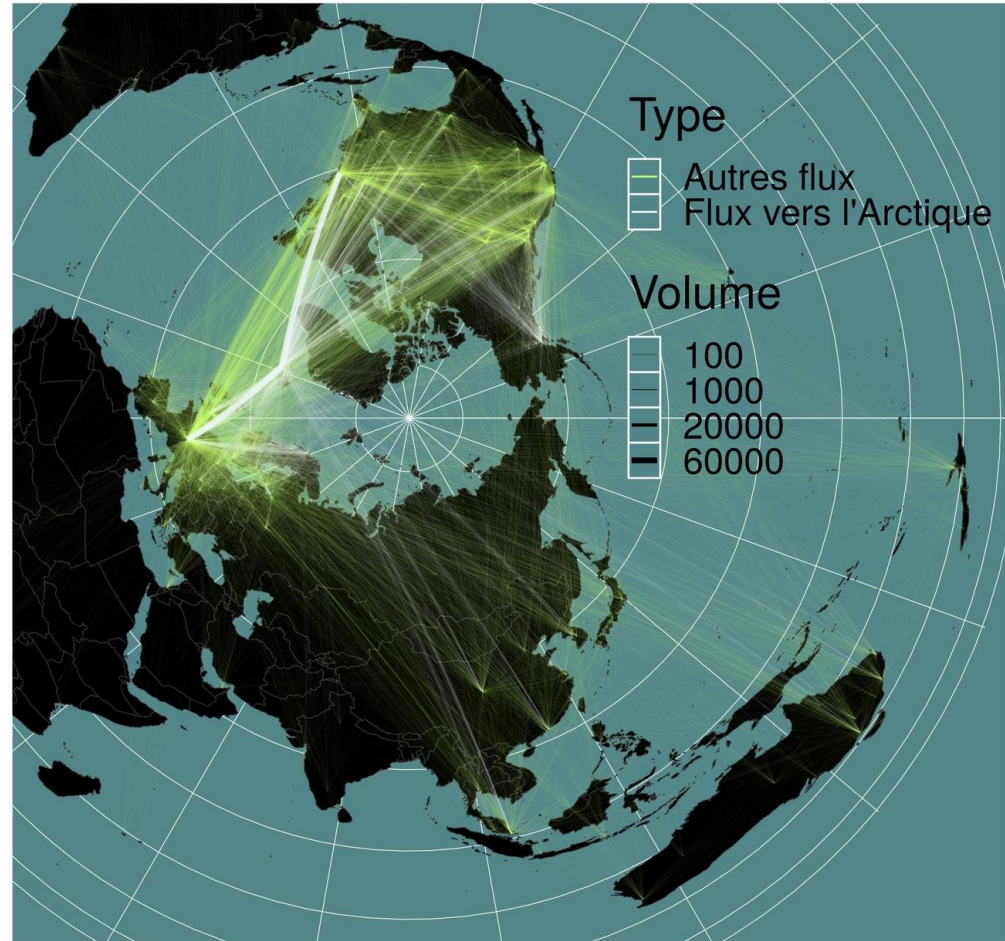
- ~24K offres de locations (coordonnées précises)
- visitées par 537057 personnes (domicile déclaratif)
- qui ont laissé 4 millions de commentaires
- chez 634453 personnes (domicile déclaratif)

Airbnb

Création de réseau :

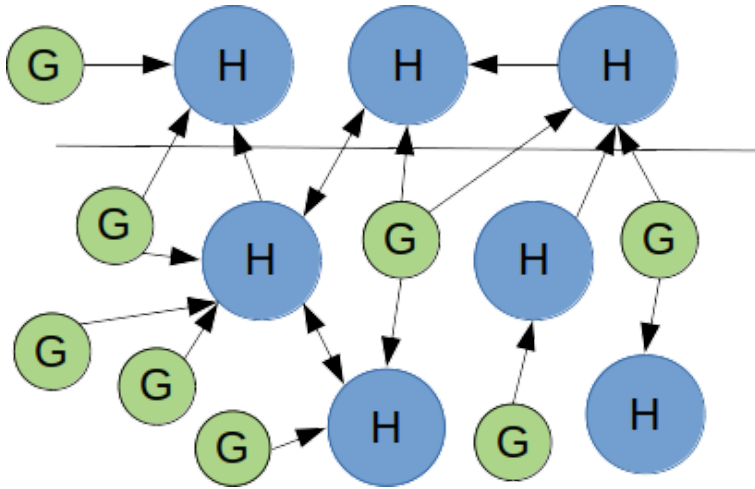


- ~24K offres de locations (coordonnées précises)
- visitées par 537057 personnes (domicile déclaratif)
- qui ont laissé 4 millions de commentaires
- chez 634453 personnes (domicile déclaratif)

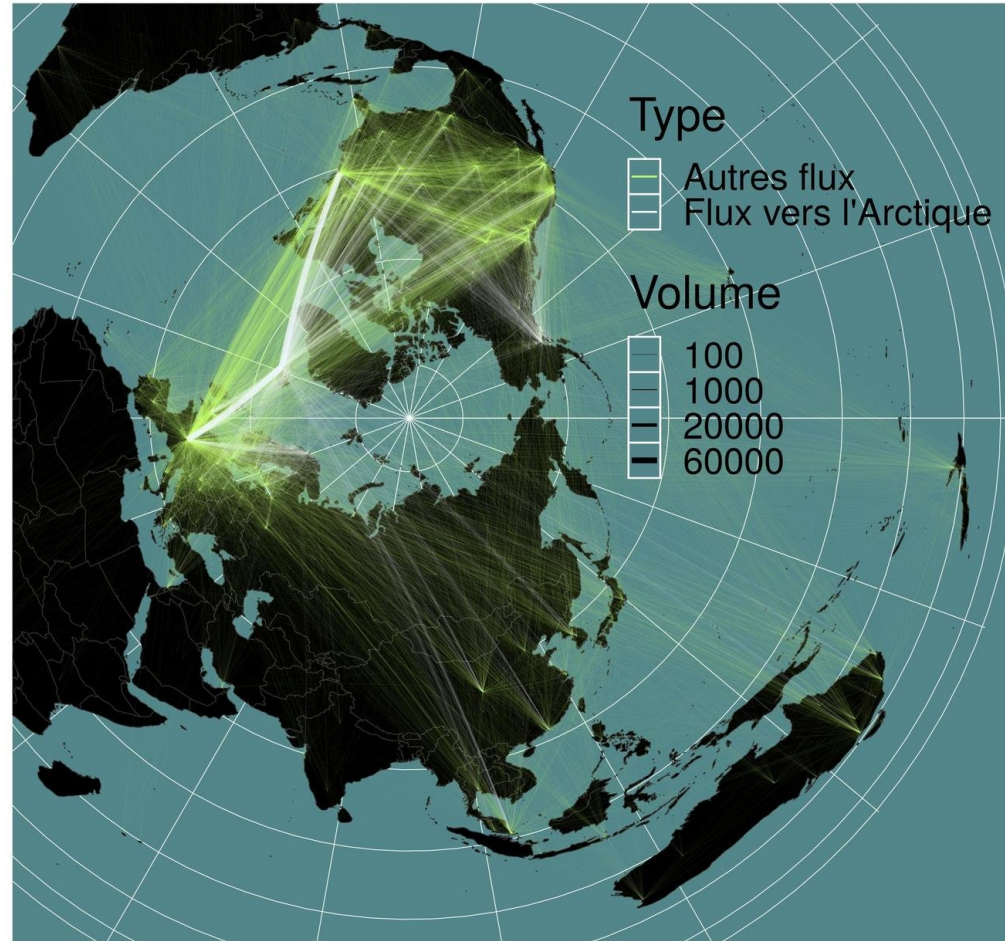


Airbnb

Création de réseau :



- ~24K offres de locations (coordonnées précises)
- visitées par 537057 personnes (domicile déclaratif)
- qui ont laissé 4 millions de commentaires
- chez 634453 personnes (domicile déclaratif)



Durée de la collecte : plusieurs semaines

Airbnb

Stayed at Horse Breeding Farm Jadar



August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008 ·

Airbnb

Stayed at Horse Breeding Farm Jadar

August 2016



Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>

110

Hi, I'm Joe

Joined in 2008



I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.

I love being a host on Airbnb and meeting great people!

Lives in San Francisco, CA

Work: Designer

Airbnb

Stayed at Horse Breeding Farm Jadar

August 2016



Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>

110

Hi, I'm Joe

Joined in 2008

“

I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.

I love being a host on Airbnb and meeting great people!

 Lives in San Francisco, CA

 Work: Designer



Stayed at Horse Breeding Farm Jaðar

August 2016



Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>

Type

Privately held company

Industry

Lodging

Founded

August 2008;
11 years ago in San Francisco, California

Founders

Brian Chesky
Joe Gebbia
Nathan Blecharczyk

Hi, I'm Joe

Joined in 2008



I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.

I love being a host on Airbnb and meeting great people!

Lives in San Francisco, CA

Work: Designer



Type

Privately held company

Industry

Lodging

Founded

August 2008; 11 years ago in San Francisco, California

Founders

Brian Chesky
Joe Gebbia
Nathan Blecharczyk

<https://www.airbnb.com/users/show/3>

Hi, I'm Brian

Joined in 2008

<https://www.airbnb.com/users/show/1>

Hi, I'm Nathan

Joined in 2008

Stayed at Horse Breeding Farm Jaðar

August 2016



Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!



Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>

Hi, I'm Joe

Joined in 2008



I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.


I love being a host on Airbnb and meeting great people!

Lives in San Francisco, CA

Work: Designer

Stayed at Horse Breeding Farm Jaðar
August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!

 Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>



airbnb

Type Privately held company

Industry Lodging

Founded August 2008;
11 years ago in San Francisco, California

Founders Brian Chesky
Joe Gebbia
Nathan Blecharczyk

<https://www.airbnb.com/users/show/3>

Hi, I'm Brian

Joined in 2008

<https://www.airbnb.com/users/show/1>

Hi, I'm Nathan

Joined in 2008

Hi, I'm Joe

Joined in 2008

“

I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.

I love being a host on Airbnb and meeting great people!

 Lives in San Francisco, CA

 Work: Designer

Joe Gebbia

From Wikipedia, the free encyclopedia

Joseph Gebbia Jr. (born August 21, 1981) is an American **billionaire** designer and Internet entrepreneur. He is the co-founder and chief product officer (CPO) of **Airbnb**. In 2009, Gebbia was listed in BusinessWeek's Top 20 Best Young Tech Entrepreneurs.^[2] In 2010, he was named in Inc. Magazine's Thirty under Thirty, and 2013, he was named in Fortune Magazine's Forty-under-Forty.^[3]



Type

Privately held company

Industry

Lodging

Founded

August 2008; 11 years ago in San Francisco, California

Founders

Brian Chesky
Joe Gebbia
Nathan Blecharczyk

<https://www.airbnb.com/users/show/3>

Hi, I'm Brian

Joined in 2008

<https://www.airbnb.com/users/show/1>


Hi, I'm Nathan

Joined in 2008

Stayed at Horse Breeding Farm Jaðar

August 2016

Great hosts! Seriously enjoyed staying here for a night. Wish to have stayed longer. They were great about late arrival, and very communicative on the Airbnb app. The farm is absolutely stunning. You wake up to an endless view over fields of horses, going almost all the way to the ocean. We got a tour of the farm and time with their awesome Icelandic dog. Highly recommend!

 Joe, San Francisco, CA
Joined in 2008

<https://www.airbnb.com/users/show/2>

Hi, I'm Joe

Joined in 2008

“

I'm a designer and athlete who has a knack for catching things when they fall. I have an eye for detail so close that I spot hairs in food. I like designing web sites and thinking about how to make interactive experiences fun and memorable.

The designers Charles and Ray Eames are my heros.

I love being a host on Airbnb and meeting great people!

 Lives in San Francisco, CA

 Work: Designer

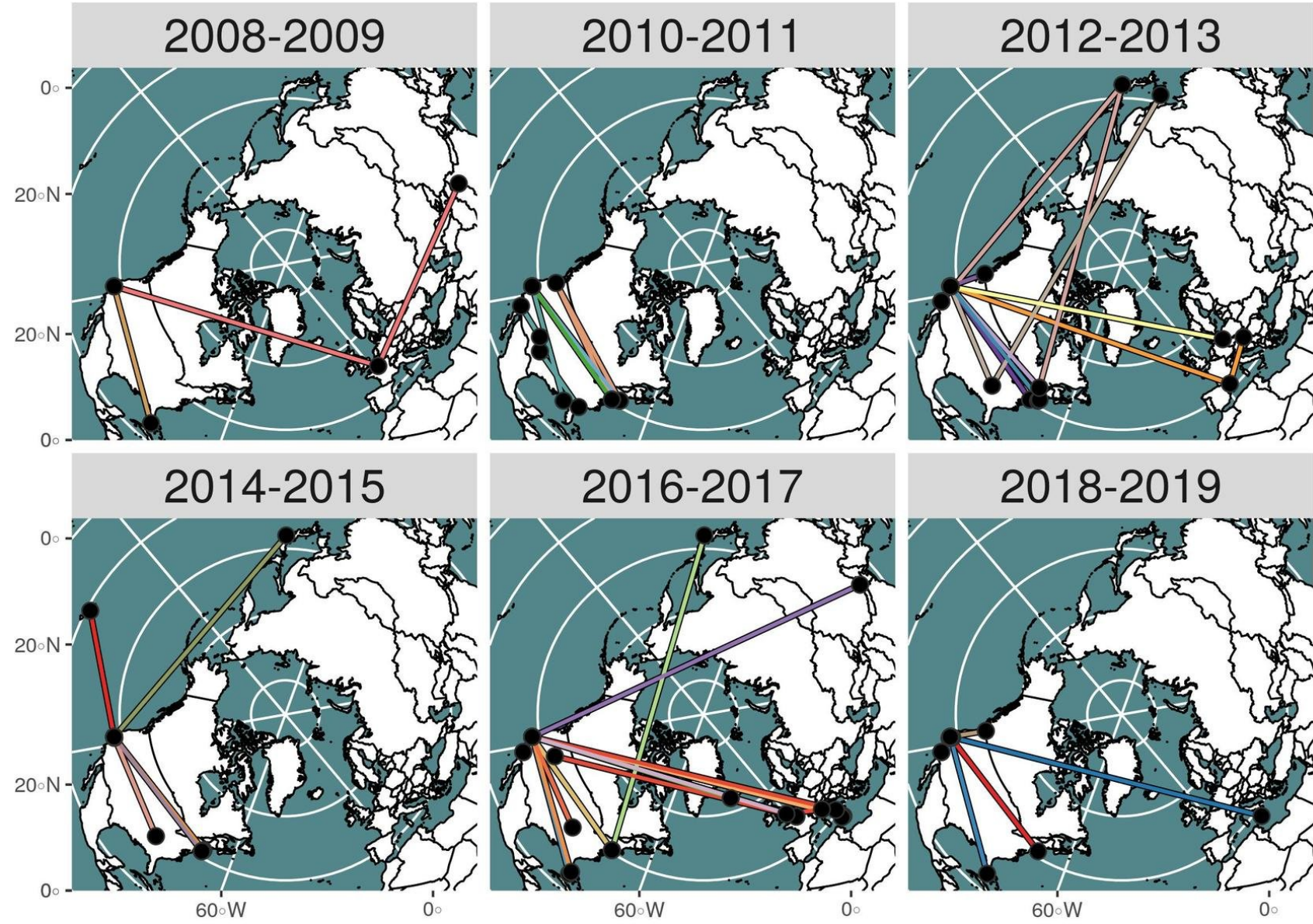
Joe Gebbia

From Wikipedia, the free encyclopedia

Joseph Gebbia Jr. (born August 21, 1981) is an American billionaire designer and Internet entrepreneur. He is the co-founder and chief product officer (CPO) of Airbnb. In 2009, Gebbia was listed in BusinessWeek's Top 20 Best Young Tech Entrepreneurs.^[2] In 2010, he was named in Inc. Magazine's Thirty under Thirty, and 2013, he was named in Fortune Magazine's Forty-under-Forty.^[3]

- Hey Joe!? Where are you going (...)?
- I'm going downtown (...)
- Uh, hey Joe, I heard you (...)

Airbnb



Mobilités urbaines



FP7 DENFREE & ANR MO3 / Eric Daudé
Modéliser et simuler les épidémies de dengue

Un des objectifs :

Comprendre le rôle des mobilités quotidiennes dans la propagation des épidémies à Bangkok & Delhi

Moyen :

Créer un modèle de mobilité, individu centré, à base d'agents (couplé à d'autres modèles)

Concept d'espace d'activité

Un des enjeux :

Collecte de données de mobilité

Mobilités urbaines



FP7 DENFREE & ANR MO3 / Eric Daudé
Modéliser et simuler les épidémies de dengue

Un des objectifs :

Comprendre le rôle des mobilités quotidiennes dans la propagation des épidémies à Bangkok & Delhi

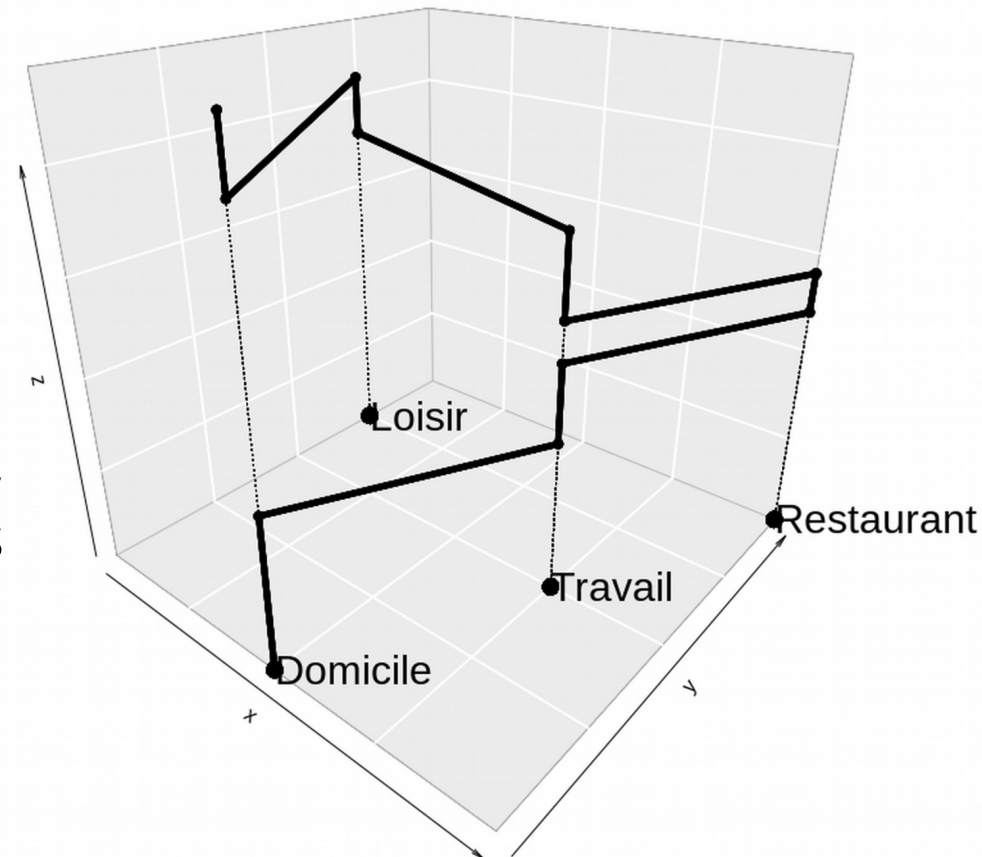
Moyen :

Créer un modèle de mobilité, individu centré, à base d'agents (couplé à d'autres modèles)

Concept d'espace d'activité

Un des enjeux :

Collecte de données de mobilité



Mobilités urbaines



FP7 DENFREE & ANR MO3 / Eric Daudé
Modéliser et simuler les épidémies de dengue

Un des objectifs :

Comprendre le rôle des mobilités quotidiennes dans la propagation des épidémies à Bangkok & Delhi

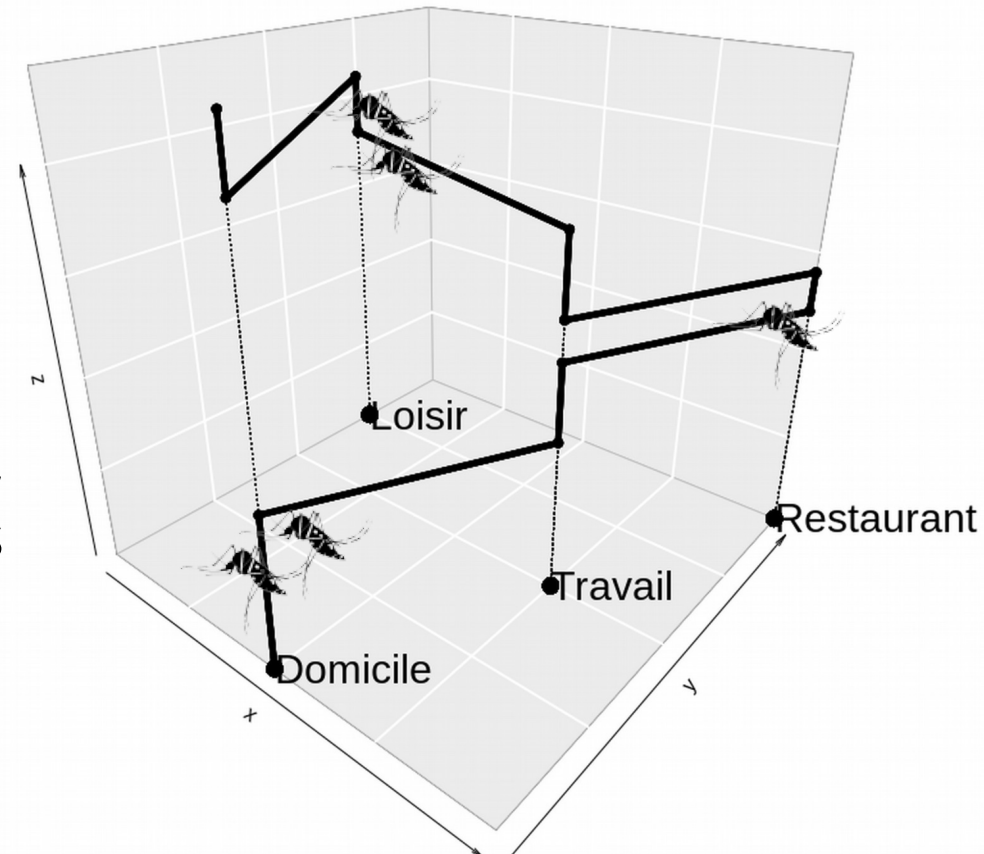
Moyen :

Créer un modèle de mobilité, individu centré, à base d'agents (couplé à d'autres modèles)

Concept d'espace d'activité

Un des enjeux :

Collecte de données de mobilité



Mobilités urbaines



FP7 DENFREE & ANR MO3 / Eric Daudé
Modéliser et simuler les épidémies de dengue

Un des objectifs :

Comprendre le rôle des mobilités quotidiennes dans la propagation des épidémies à Bangkok & Delhi

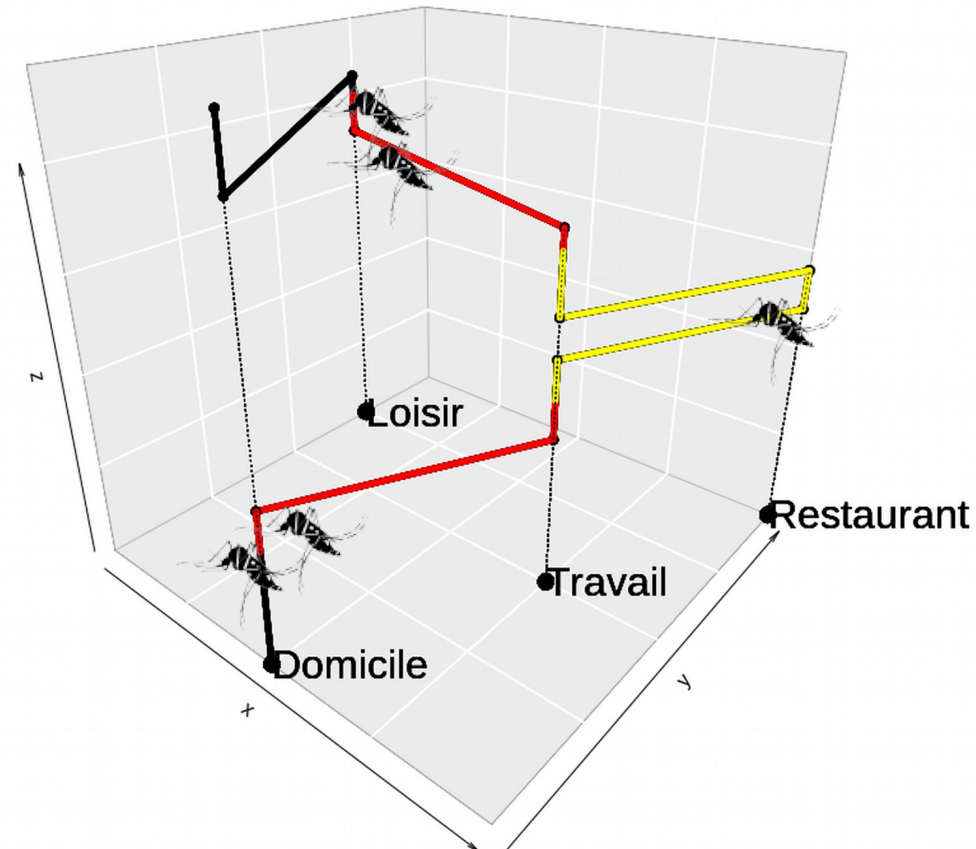
Moyen :

Créer un modèle de mobilité, individu centré, à base d'agents (couplé à d'autres modèles)

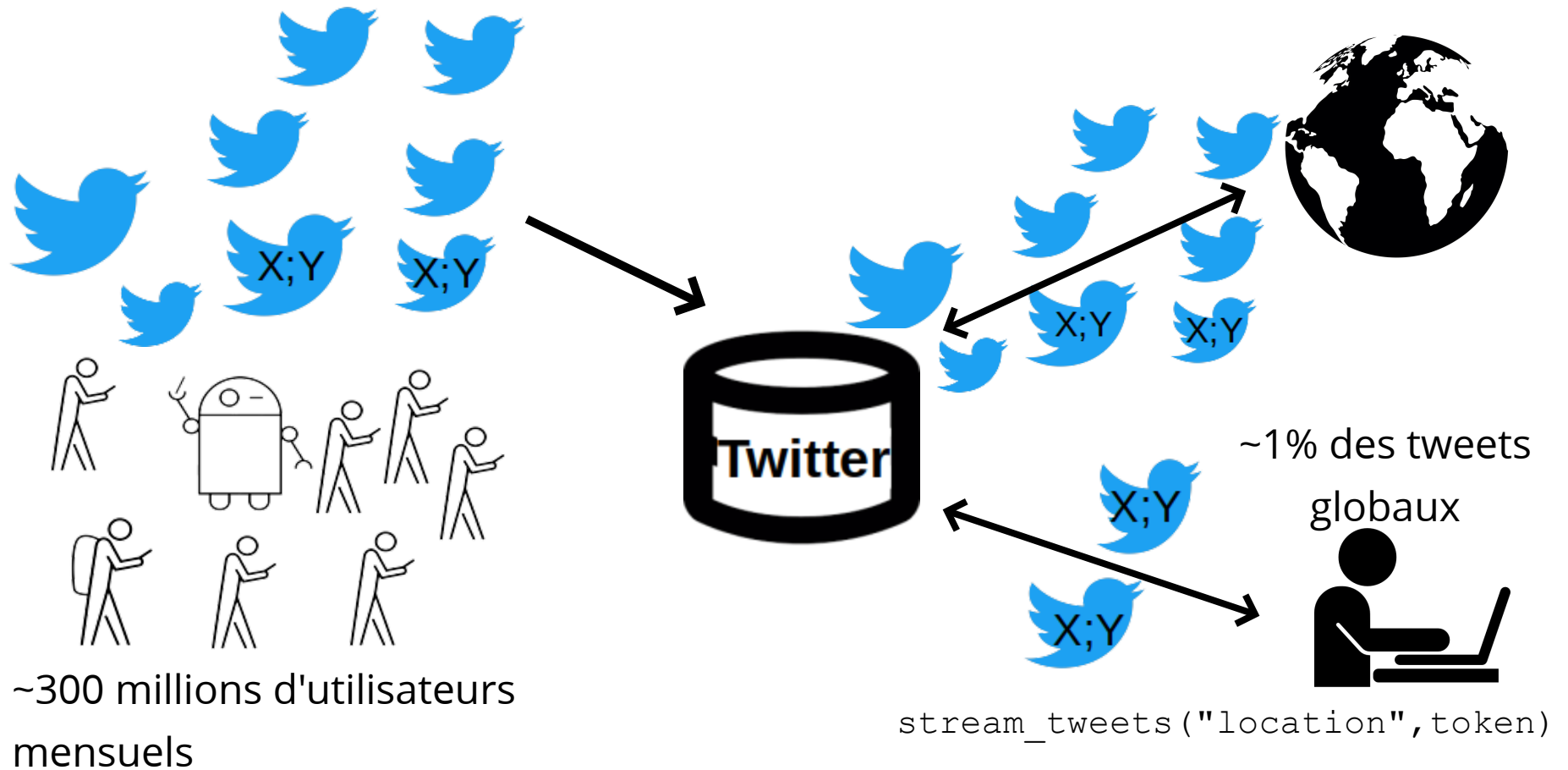
Concept d'espace d'activité

Un des enjeux :

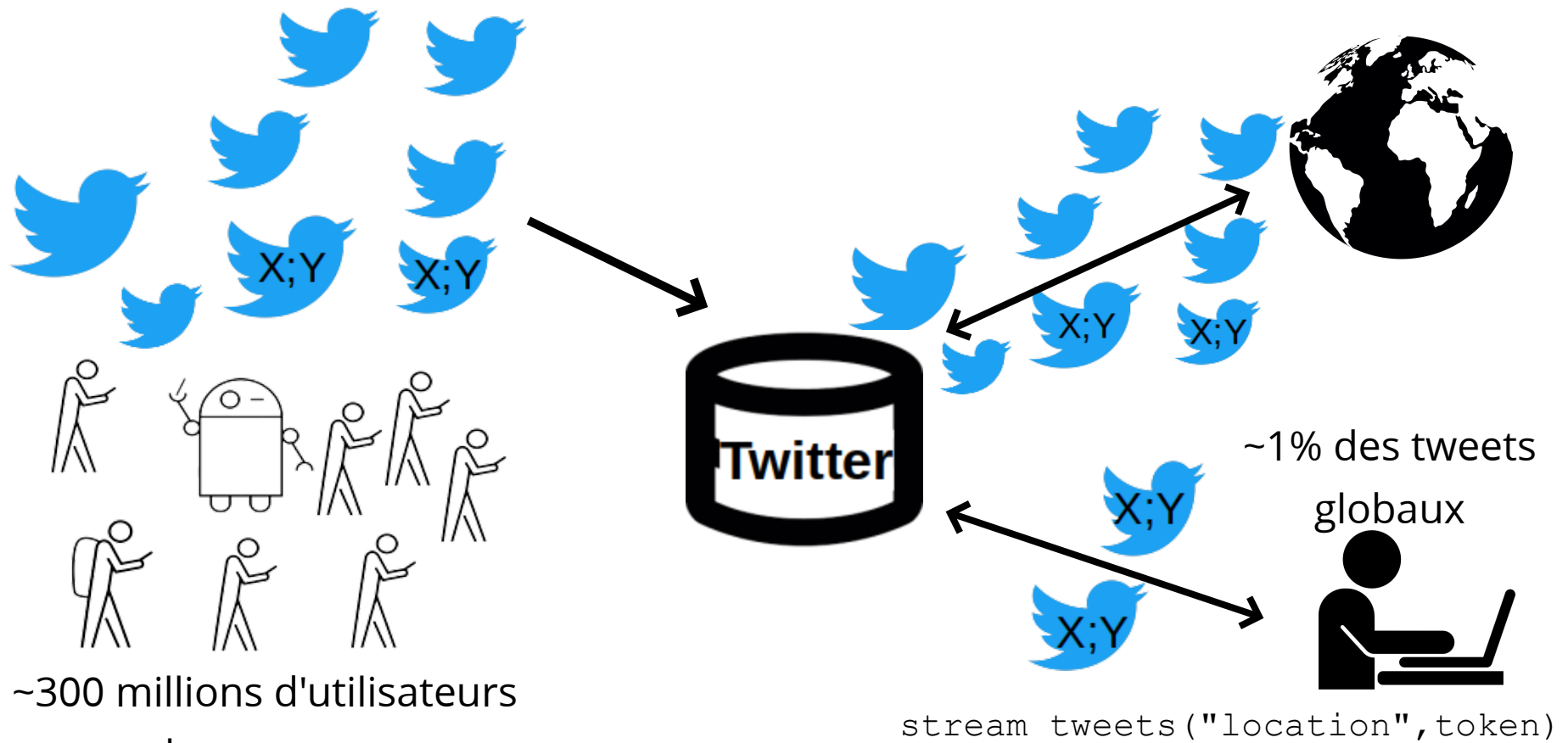
Collecte de données de mobilité



Twitter

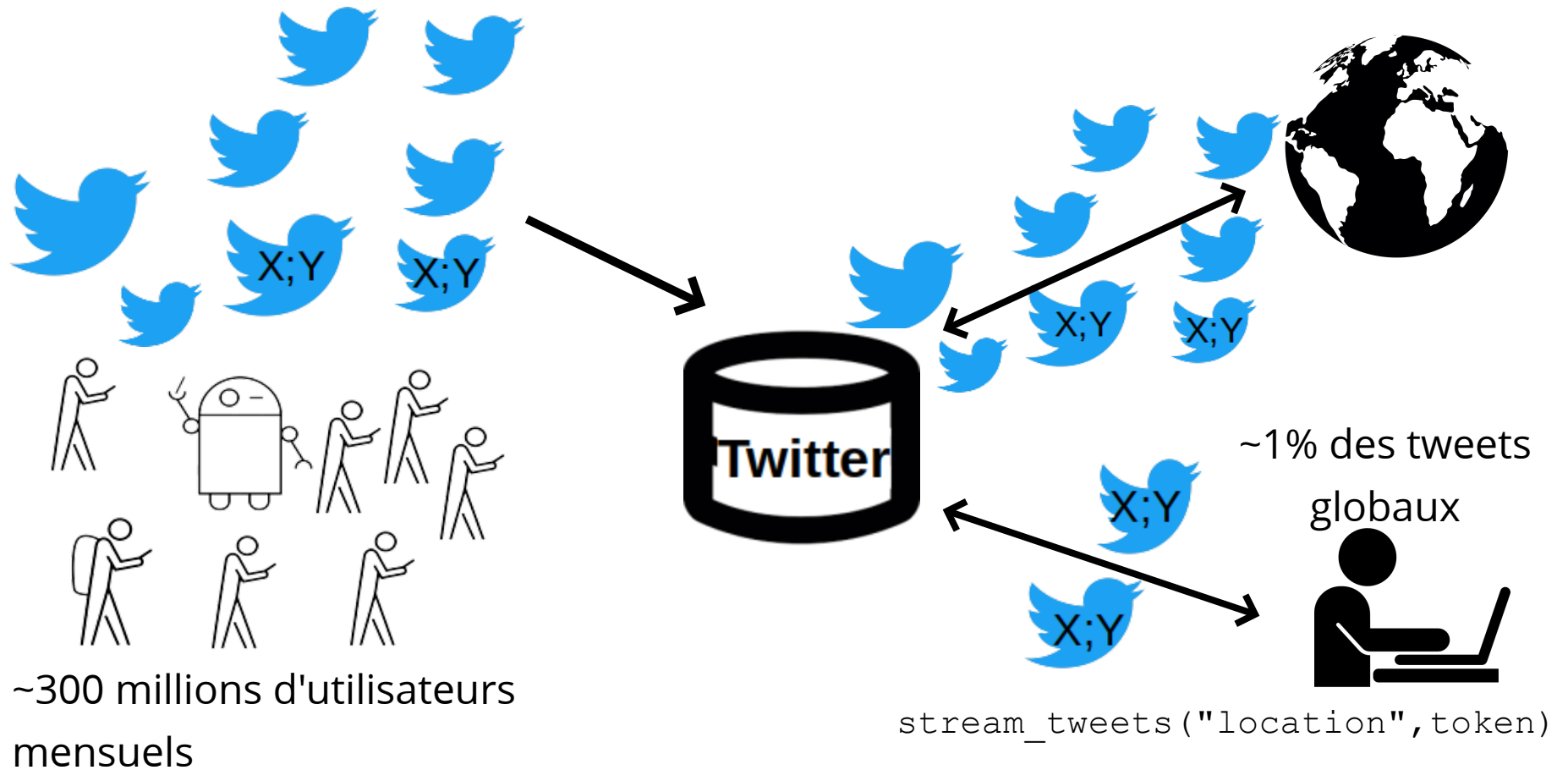


Twitter

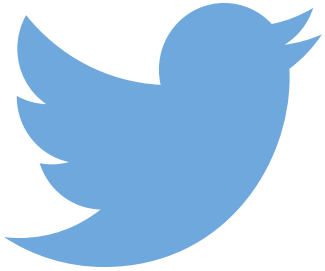


Juin 2014 – Décembre 2015 : Bangkok (~10M Hab.)
Données brutes ~ 30M de tweets / ~300K utilisateurs
Données filtrées ⇒ 17M / ~76K utilisateurs

Twitter



Twitter



Collecter des données de mobilités

Site web : Twitter

Langage : Python ou R

Framework : Scrapy

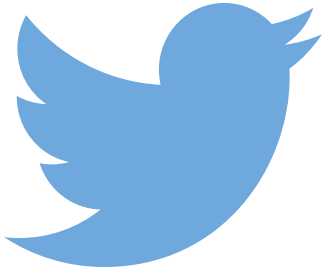
Type de récolte : Continue

Format : csv

Difficultés :

- Mon premier code sous Python
- Nécessite un jeton (API)
- Accès a un échantillon
- Stockage et traitement des données

Twitter



Collecter des données de mobilités

Site web : Twitter

Langage : Python ou R

Framework : Scrapy

Type de récolte : Continue

Format : csv

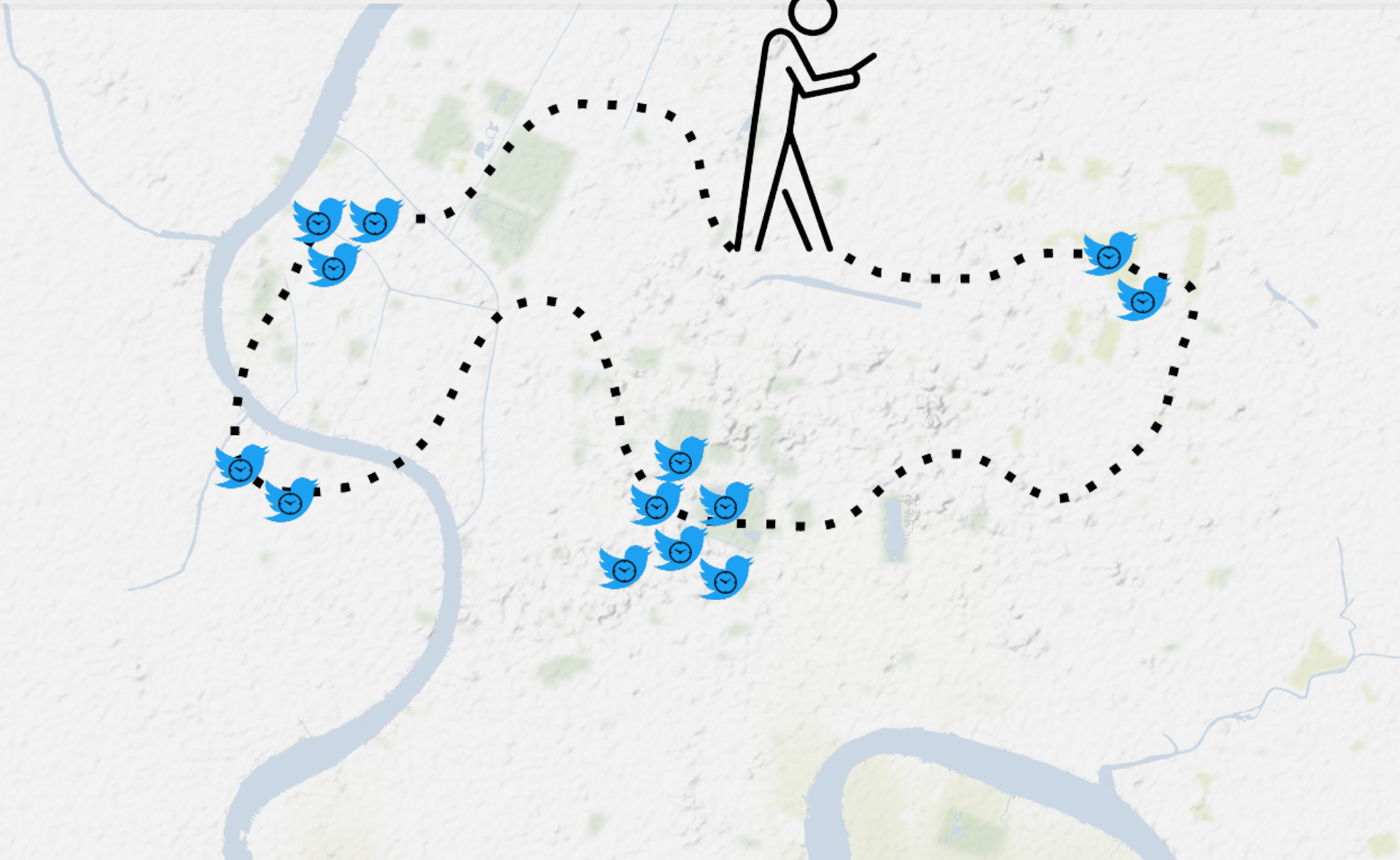
Difficultés :

- Mon premier code sous Python
- Nécessite un jeton (API)
- Accès a un échantillon
- Stockage et traitement des données

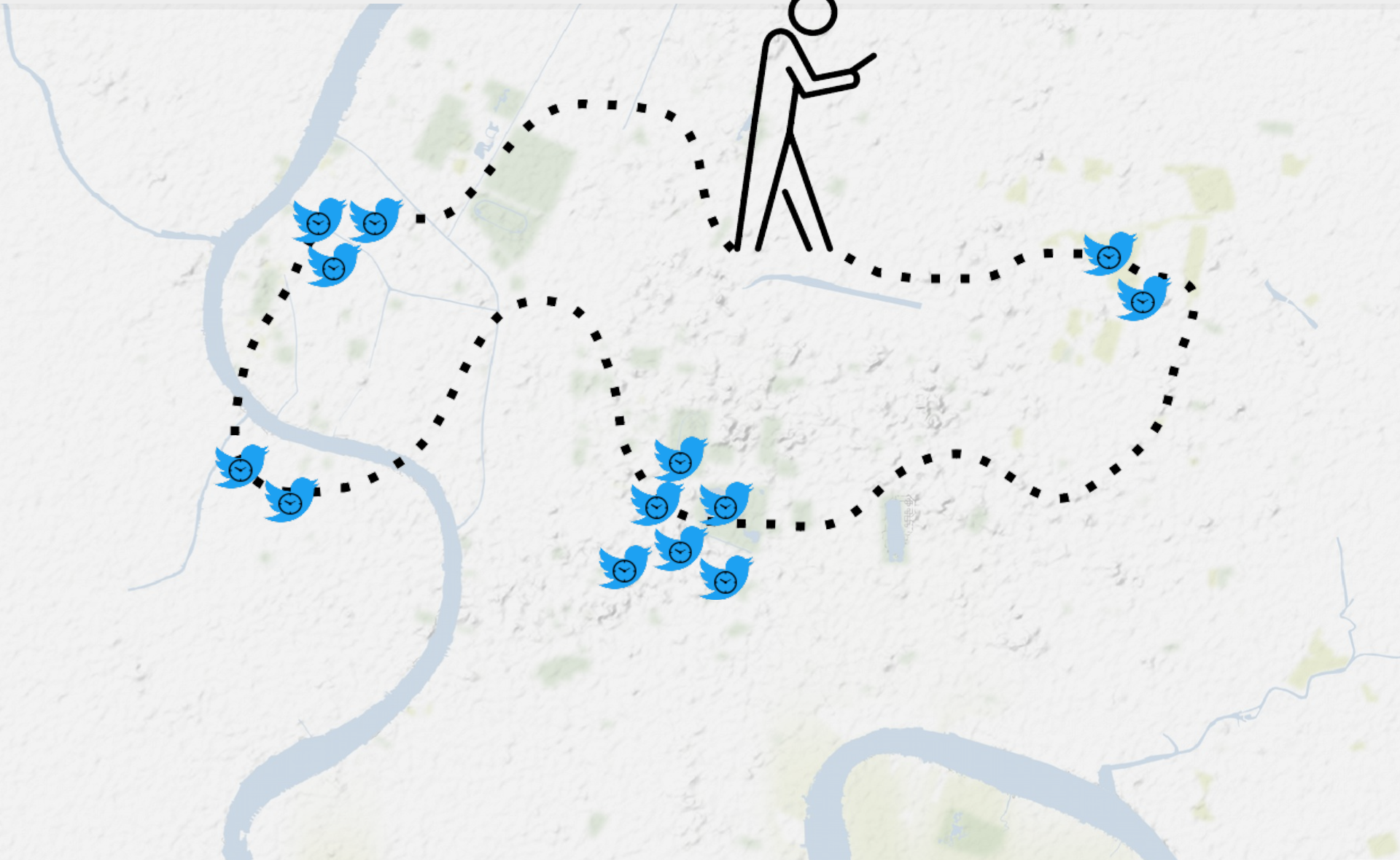
Bangkok (06/2014 – 12/2015)

- Données brutes :
 - ~ 30M de tweets
 - ~300K utilisateurs
- Données filtrées :
 - ~ 17M de tweets
 - ~ 76K utilisateurs

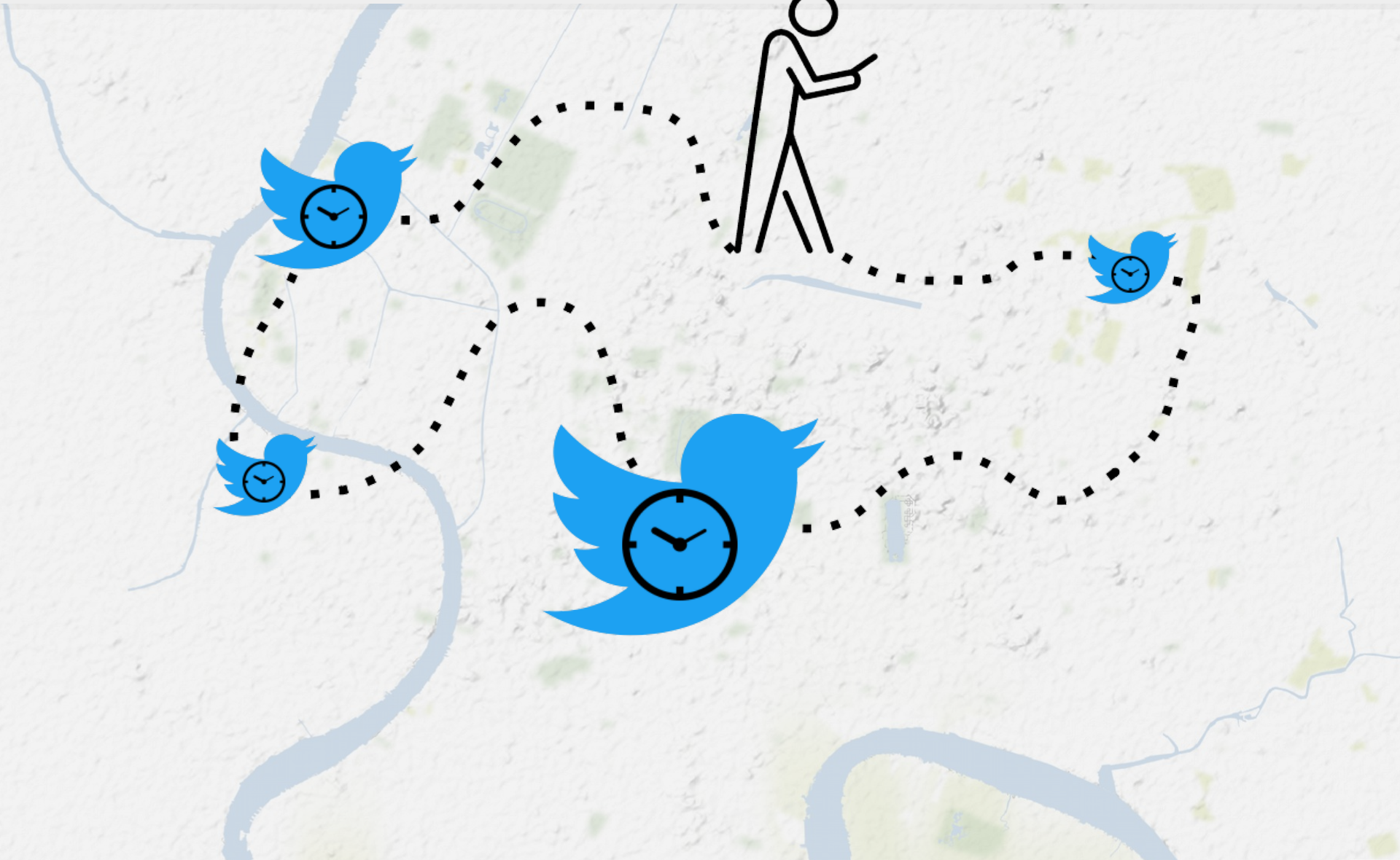
Twitter & géolocalisation



Twitter & géolocalisation

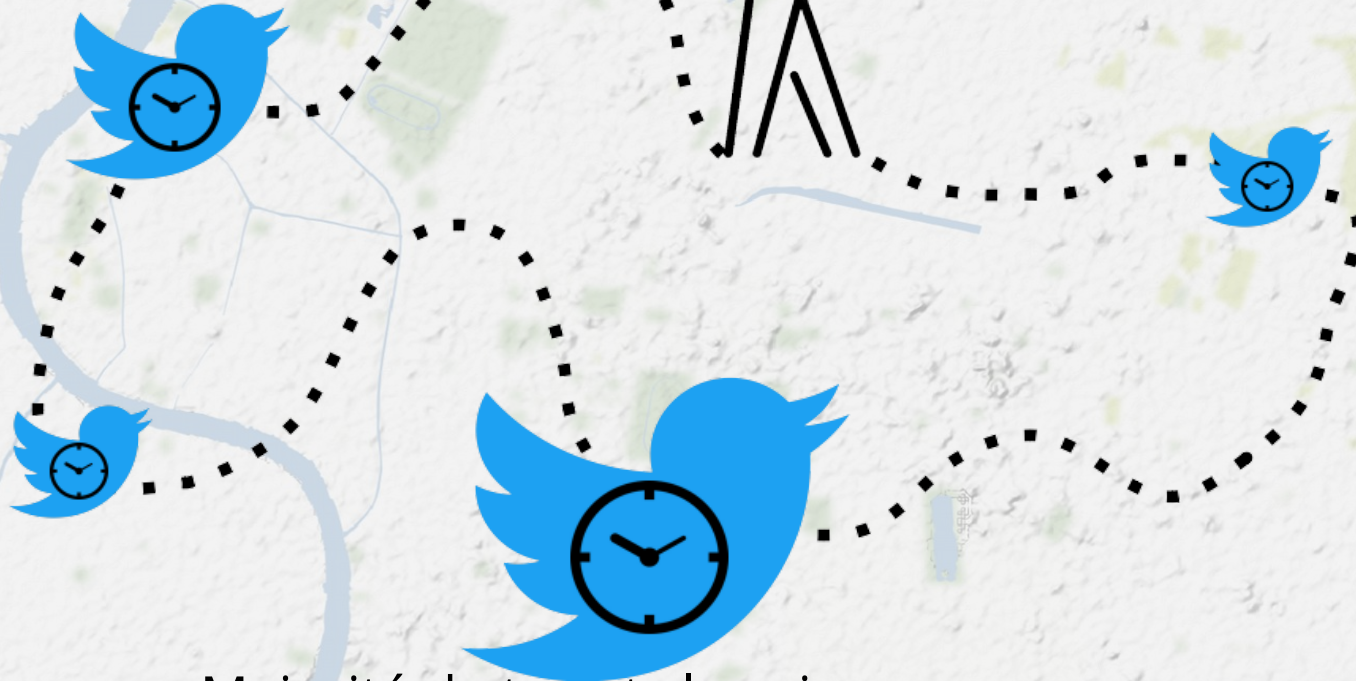


Twitter & géolocalisation



Twitter & géolocalisation

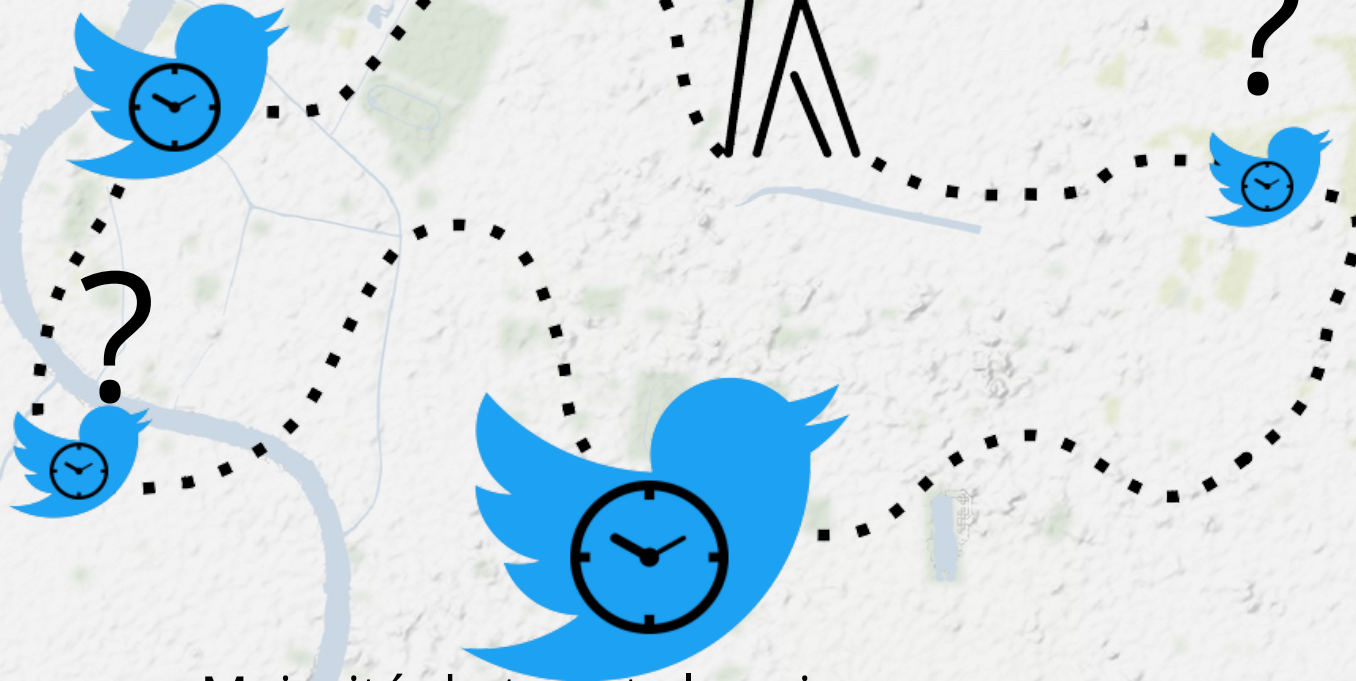
Majorité de tweets en journée
==> Lieux de travail



Majorité de tweets le soir
==> Domicile

Twitter & géolocalisation

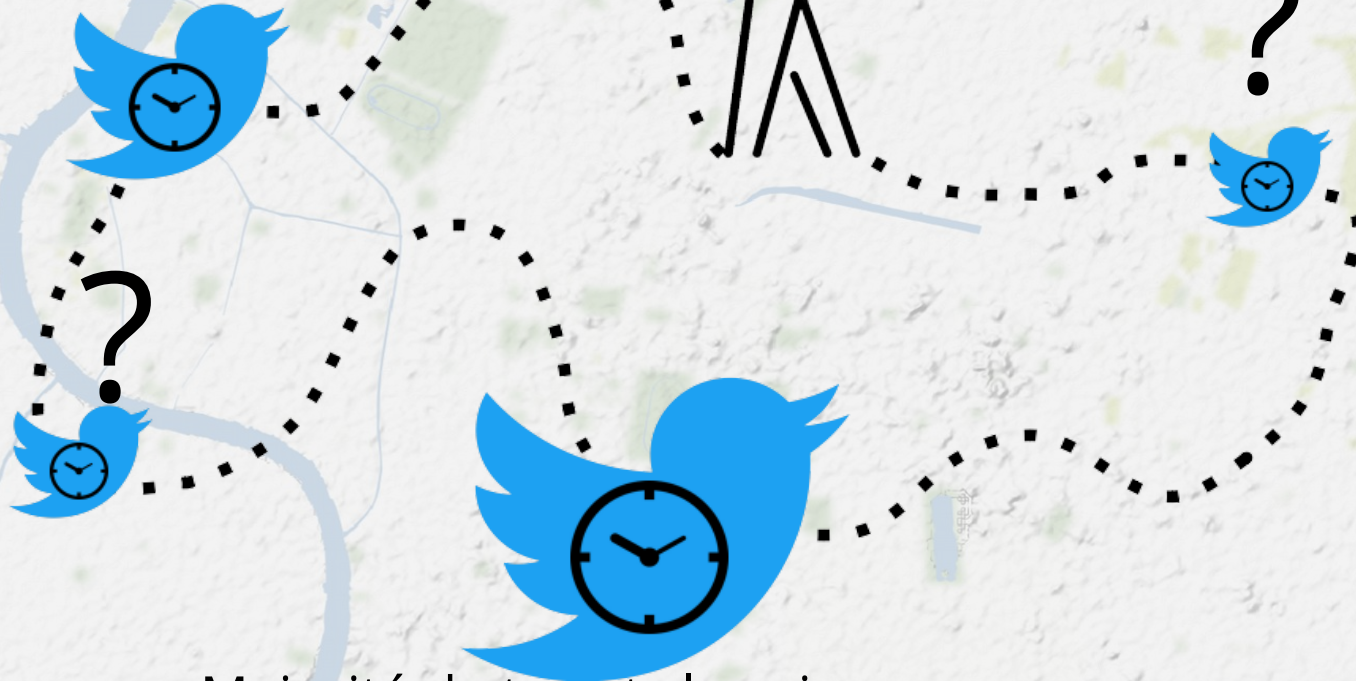
Majorité de tweets en journée
==> Lieux de travail



Majorité de tweets le soir
==> Domicile

Twitter & géolocalisation

Majorité de tweets en journée
==> Lieux de travail

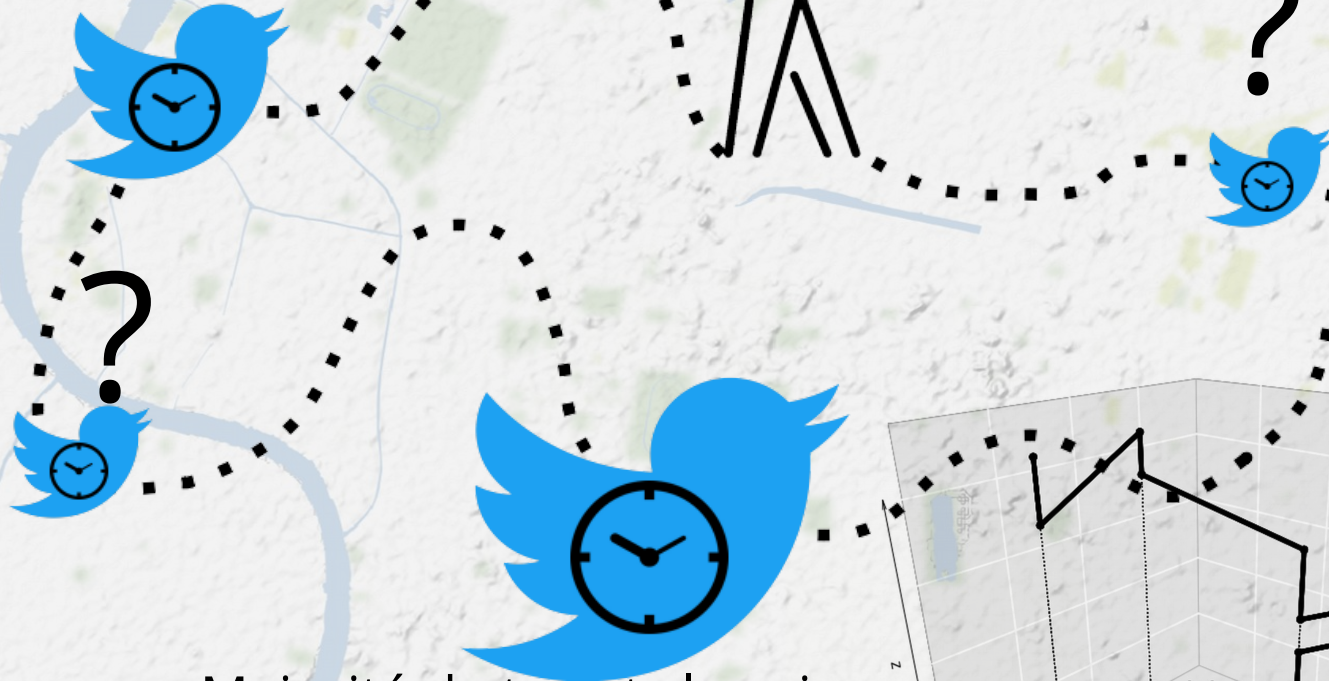


Majorité de tweets le soir
==> Domicile

Espace d'activité incomplet...

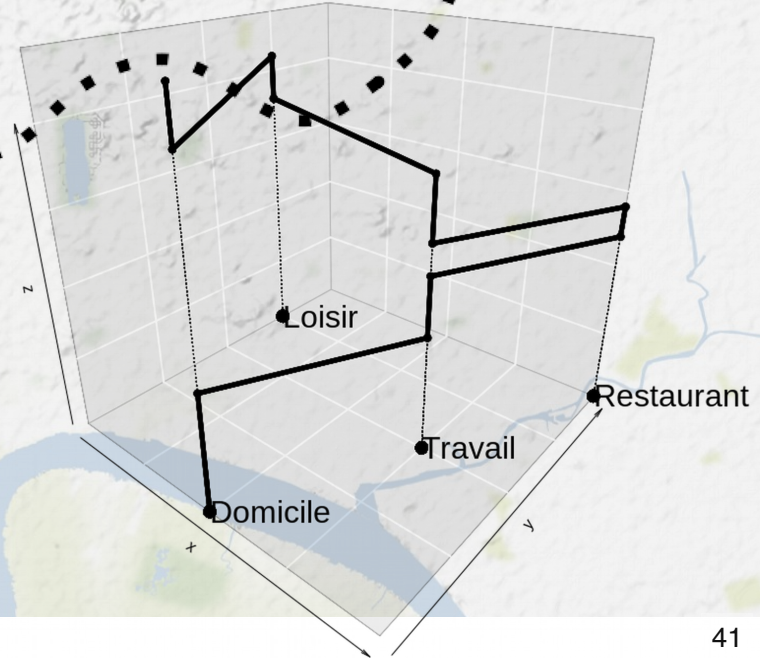
Twitter & géolocalisation

Majorité de tweets en journée
==> Lieux de travail



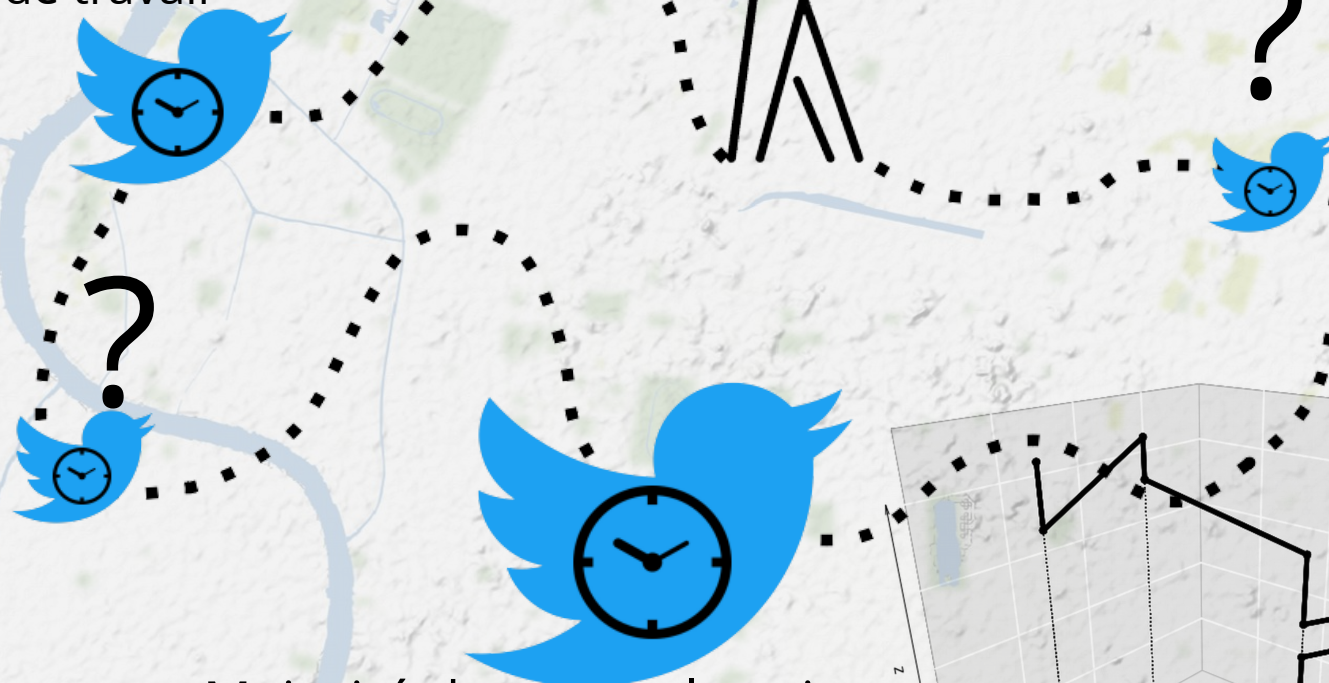
Majorité de tweets le soir
==> Domicile

Espace d'activité incomplet...



Twitter & géolocalisation

Majorité de tweets en journée
==> Lieux de travail

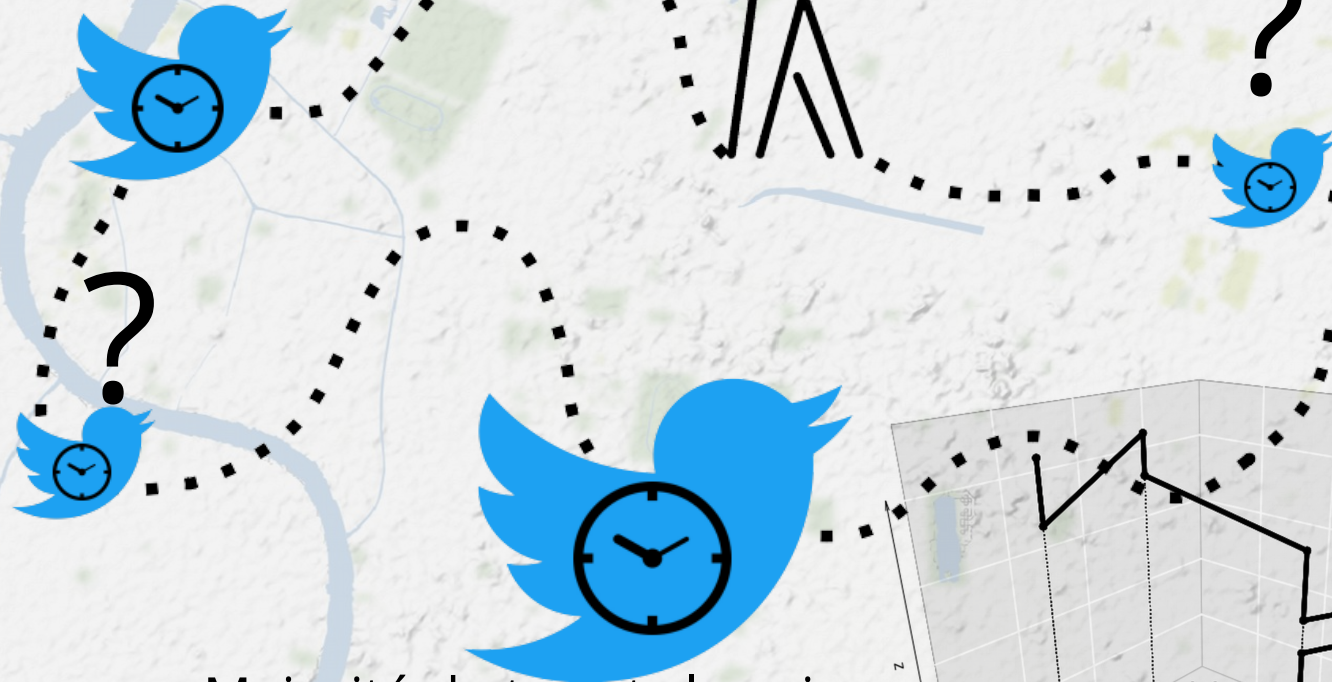


Majorité de tweets le soir
==> Domicile

Espace d'activité incomplet...

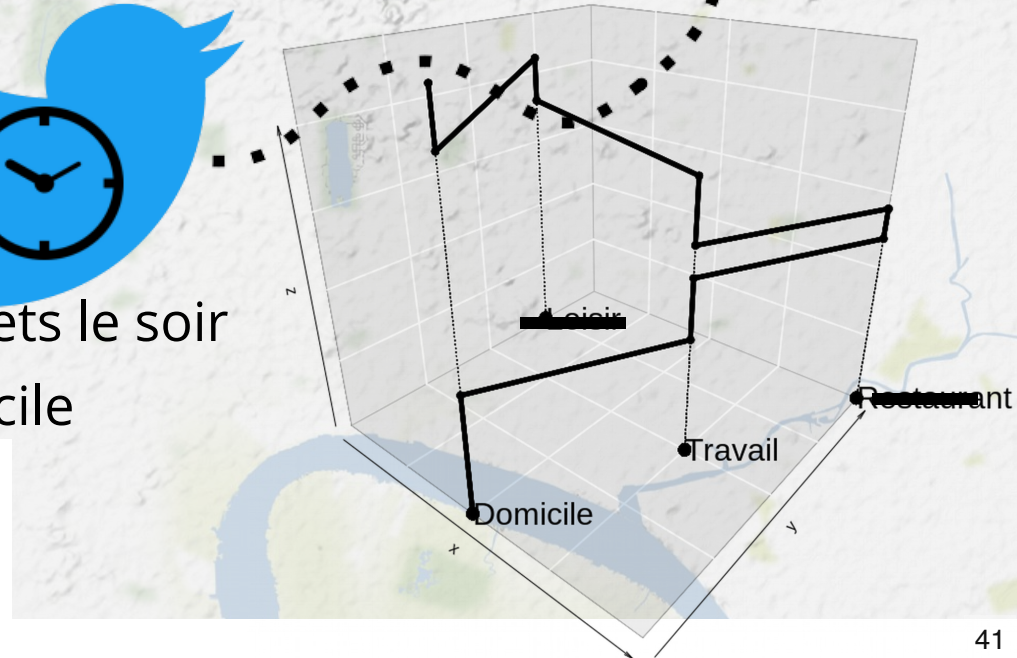
Twitter & géolocalisation

Majorité de tweets en journée
==> Lieux de travail



Majorité de tweets le soir
==> Domicile

Espace d'activité incomplet...
=> Couche d'utilisation du sol ? OSM ?
peu développé à Bangkok ...



Google maps



Collecter des informations sur l'utilisation du sol à Bangkok pour compléter l'espace d'activité Twitter

Site web : google maps

Langage : Python ou R

Framework : Scrapy / httr

Type de récolte : Campagne

Format : json

Google maps



Collecter des informations sur l'utilisation du sol à Bangkok pour compléter l'espace d'activité Twitter

Site web : google maps

Langage : Python ou R

Framework : Scrapy / httr

Type de récolte : Campagne

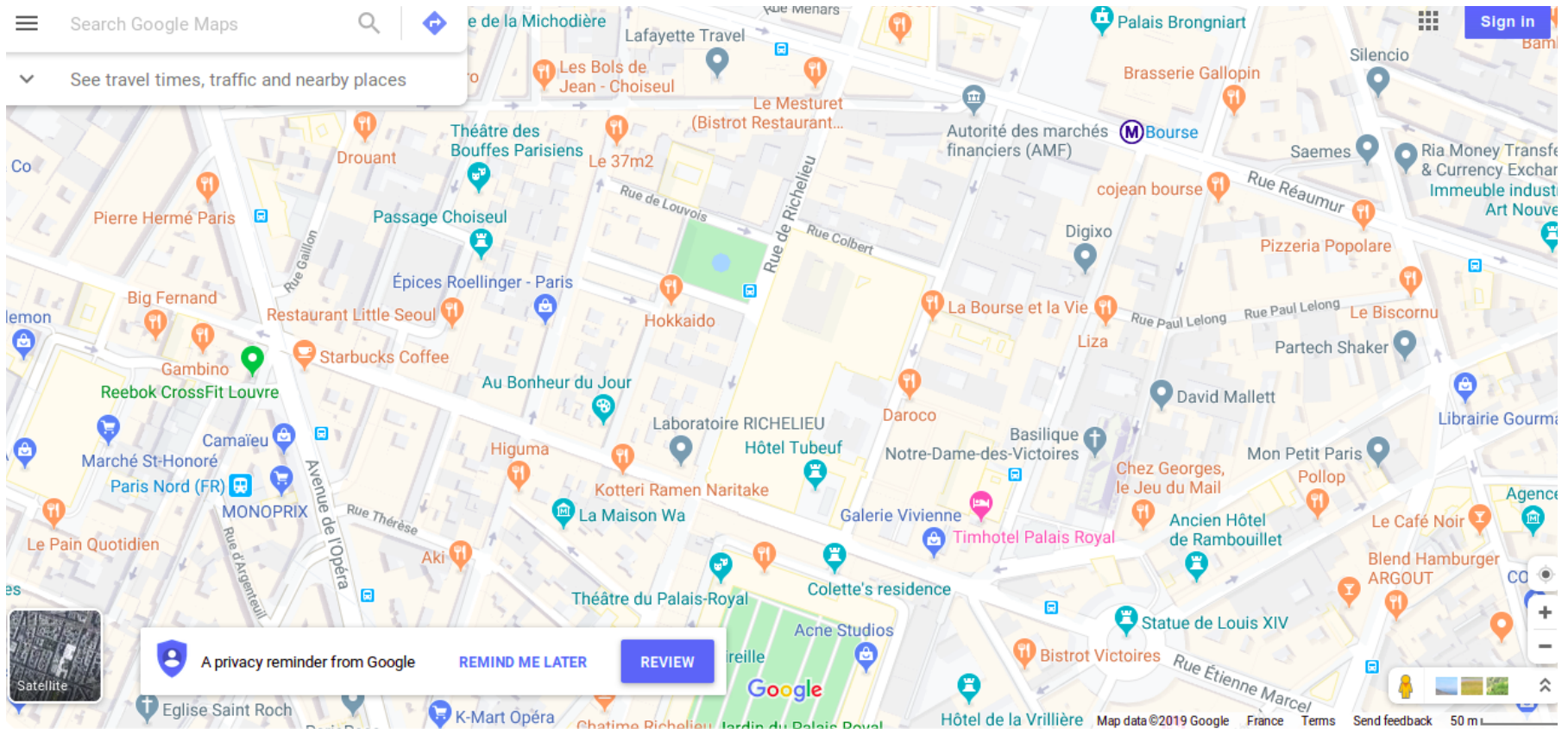
Format : json

Difficultés :

- Nécessite un jeton (API)
- Requêtes gratuites limitées
- Zone d'étude importante : Bangkok
- hiérarchisation de l'information renvoyées
- Évolution des conditions d'utilisation de l'API

Google maps

- POI : Point of Interest ⇒ Localisation et catégorie des lieux de la base



Google maps

- POI : Point of Interest ⇒ Localisation et catégorie des lieux de la base
- Fonction « nearby search » de l'API Place

[https://maps.googleapis.com/maps/api/place/nearbysearch/json?](https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=LAT,LON&types=*&radius=100&key=TOKEN) + Fenêtre
location=**LAT,LON**&types=*&**radius**=100&key=**TOKEN** mobile

Google maps

- POI : Point of Interest ⇒ Localisation et catégorie des lieux de la base
- Fonction « nearby search » de l'API Place

<https://maps.googleapis.com/maps/api/place/nearbysearch/json?> + Fenêtre mobile
location=**LAT,LON**&types=*&**radius**=100&key=**TOKEN**

```
geometry:
  ▼ location:
    lat: 48.8271088
    lng: 2.3829514
  ▶ viewport: {}
icon: "https://maps.gstatic.com/mapfiles/pl
id: "2943b9e5ce667218a32528f98bca810b90e5
name: "Université Paris Diderot"
place_id: "ChIJjwYqvTly5kcR2gyU-cIu3no"
plus_code: {}
rating: 4
reference: "ChIJjwYqvTly5kcR2gyU-cIu3no"
scope: "GOOGLE"
types:
  0: "university"
  1: "point_of_interest"
  2: "establishment"
user_ratings_total: 3
vicinity: "17 Rue Jean Antoine de Baïf, Paris"
```

Google maps

- POI : Point of Interest ⇒ Localisation et catégorie des lieux de la base
- Fonction « nearby search » de l'API Place

<https://maps.googleapis.com/maps/api/place/nearbysearch/json?> + Fenêtre mobile
location=**LAT,LON**&types=***&radius=100**&key=**TOKEN** mobile

```
geometry:  
  ▼ location:  
    lat: 48.8271088  
    lng: 2.3829514  
  ▶ viewport: {}  
icon: "https://maps.gstatic.com/mapfiles/pl  
id: "2943b9e5ce667218a32528f98bca810b90e5  
name: "Université Paris Diderot"  
place_id: "ChIJjwYqvTly5kcR2gyU-cIU3no"  
plus_code: {}  
rating: 4  
reference: "ChIJjwYqvTly5kcR2gyU-cIU3no"  
scope: "GOOGLE"  
types:  
  0: "university"  
  1: "point_of_interest"  
  2: "establishment"  
user_ratings_total: 3  
vicinity: "17 Rue Jean Antoine de Baïf, Paris"
```



Google maps

Administration



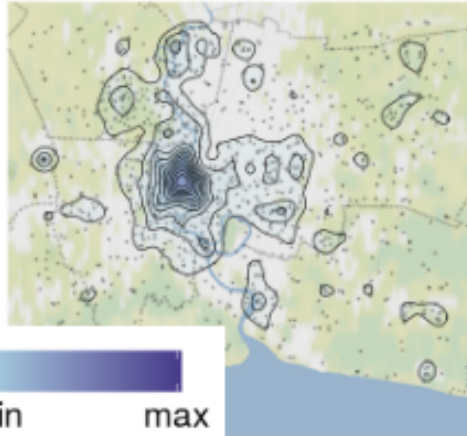
Entertainment



Hospital/Health



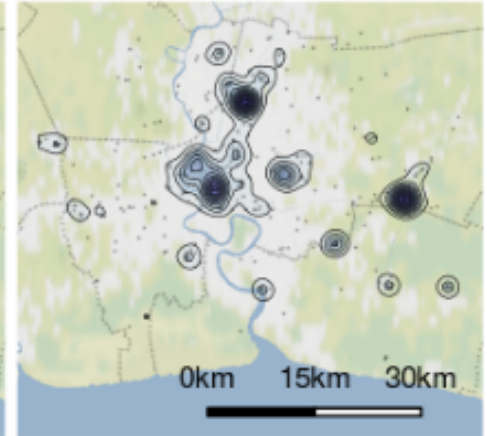
Place of Worship



School



University/College



Bangkok :
~ 100K POI
~100 catégories

Nombre limite de requêtes quotidiennes gratuites :
~2000 jusqu'en juin 2018 ; 5 aujourd'hui

Google

AOI: « Areas of Interest - places where there's a lot of activities and things to do (..) We determine areas of interest with an algorithmic process that allows us to highlight the areas with the highest concentration of restaurants, bars and shops »



API « Static Map »

⇒ Afficher le fond de carte (ggmap, R)

⇒ Enregistrer l'image

⇒ Extraire les pixels selon leur couleur

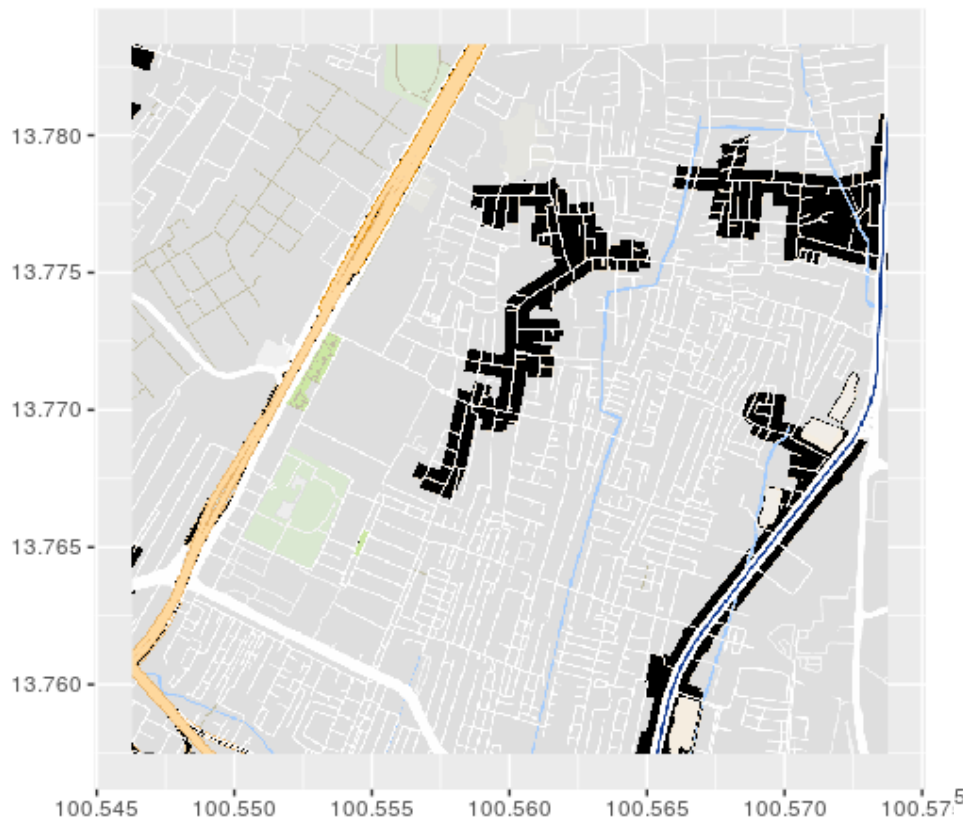
S'inspirer de la composition des AOI en POI

⇒ Générer de nouveaux AOI

⇒ Typologie

Google

AOI: « Areas of Interest - places where there's a lot of activities and things to do (..) We determine areas of interest with an algorithmic process that allows us to highlight the areas with the highest concentration of restaurants, bars and shops »



API « Static Map »

⇒ Afficher le fond de carte (ggmap, R)

⇒ Enregistrer l'image

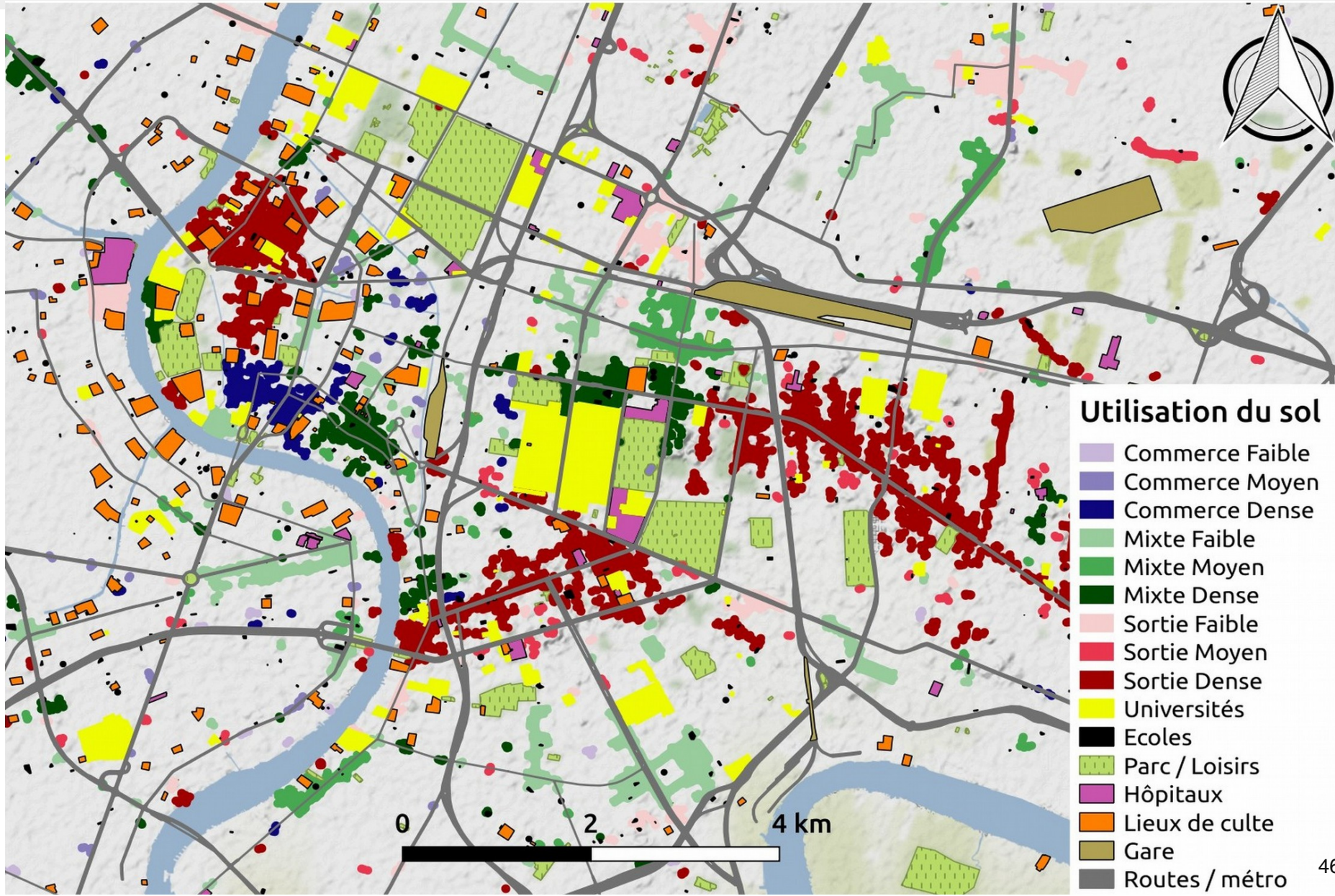
⇒ Extraire les pixels selon leur couleur

S'inspirer de la composition des AOI en POI

⇒ Générer de nouveaux AOI

⇒ Typologie

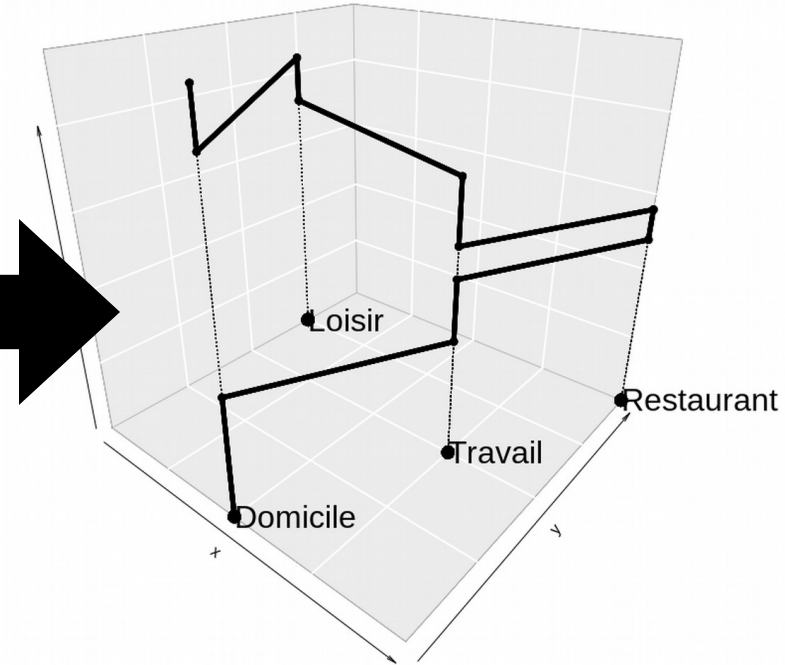
Twitter & Google maps



Twitter & Google maps



Twitter & Google



Pour chaque utilisateur de Twitter :

- Reconstruire des espaces d'activités **hypothétiques**

Coté modélisateur :

- utiliser les statistiques de visites pour générer des agents synthétiques mobiles

Twitter & Google

Quelques précautions pour améliorer le niveau d'anonymat des personnes :

- Changement du nom de l'utilisateur
- Flou temporel :
 - Mardi 10 juillet 2015 18h ==> Mardi 18h
 - Séquences de visites ==> Fréquences de visites
- Flou spatial :
 - Domicile agrégé au sous district (~arrondissement)
 - Autre lieux dans des mailles (250m)
- Flou thématique :
 - On ne sait pas exactement quelle activité réalisée ("lieu de sortie", "lieu d'éducation" etc.)

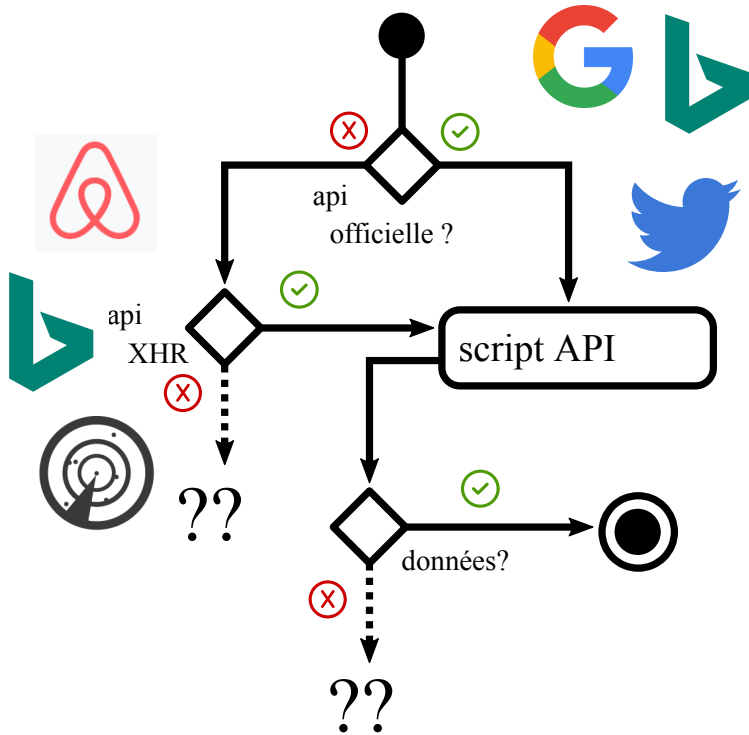
Sécurisation des données :

- Fichiers encryptés sur serveur protégé (Huma-num)

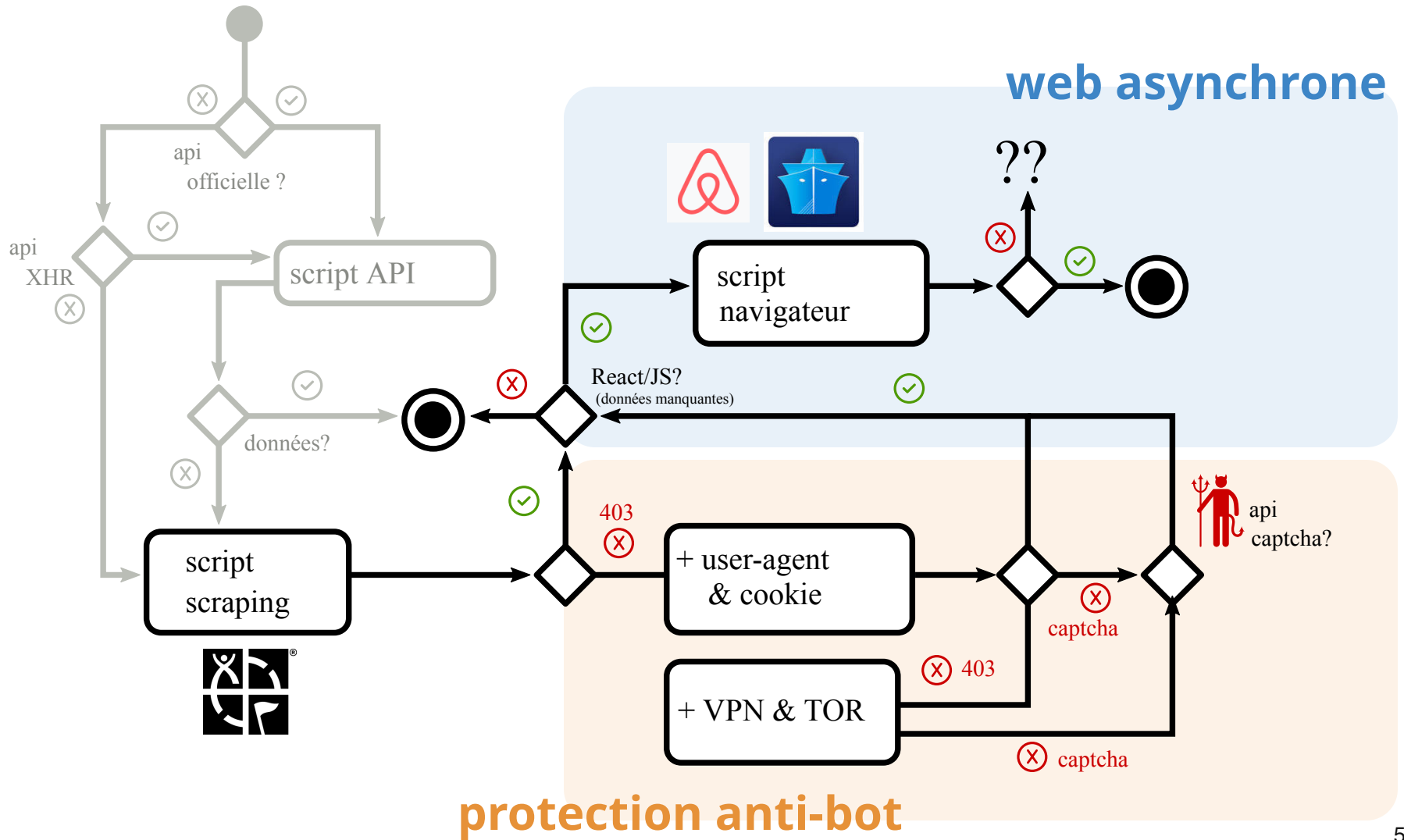
Printemps 2015 : **Mise à jour de Twitter** :

Désactivation automatique de la géolocalisation après l'envoi d'un message géolocalisé

Protocoles

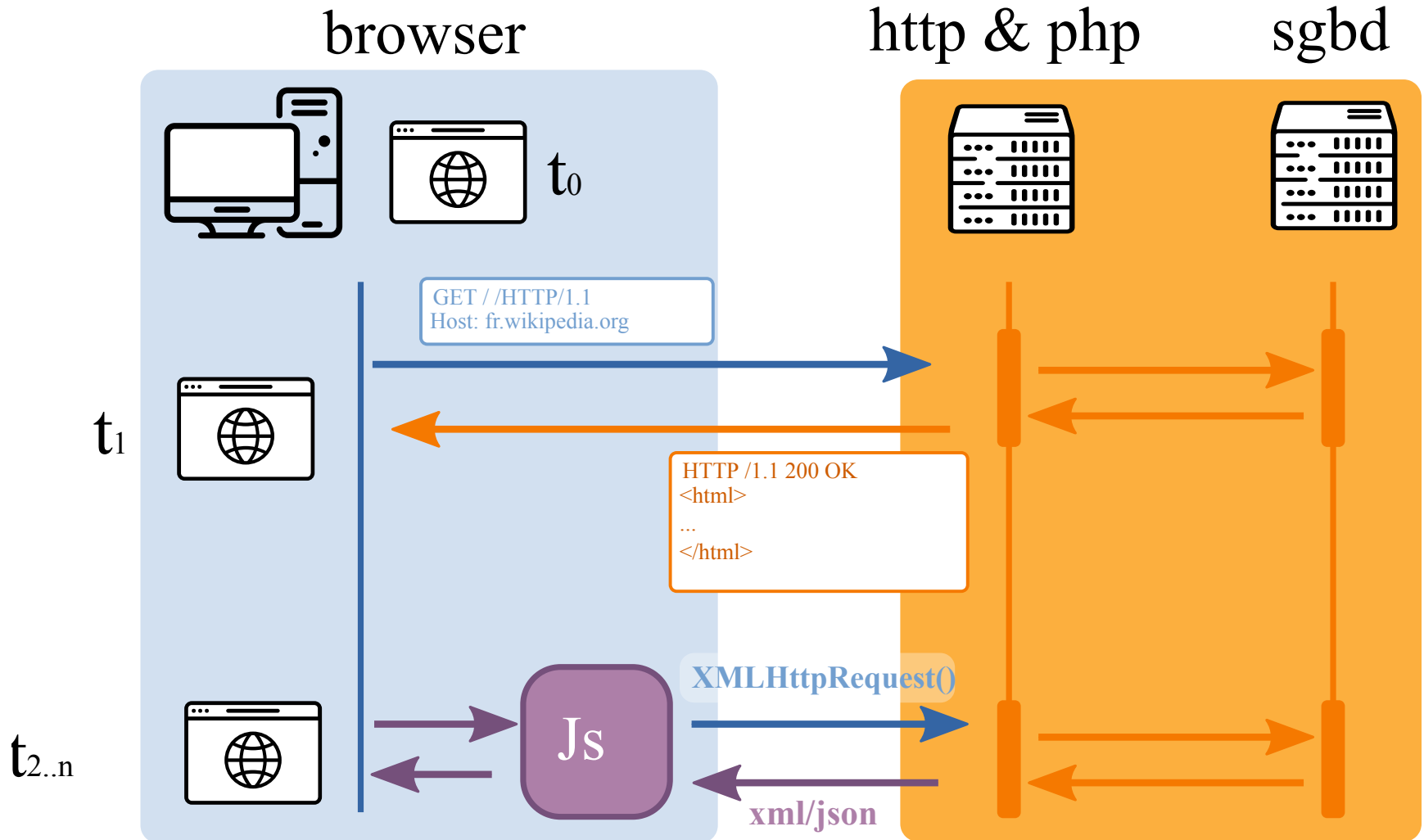


Zone de ~turbulence~



WEB asynchrone

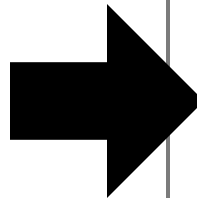
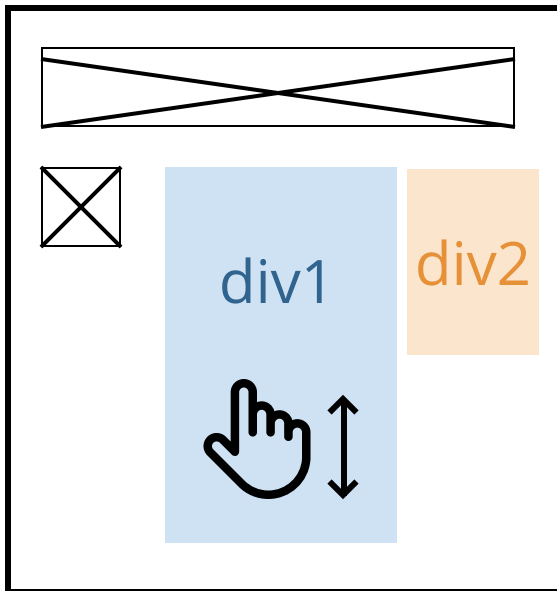
Javascript / AJAX



WEB asynchrone

Javascript + React

"What you see *is not* what you get"



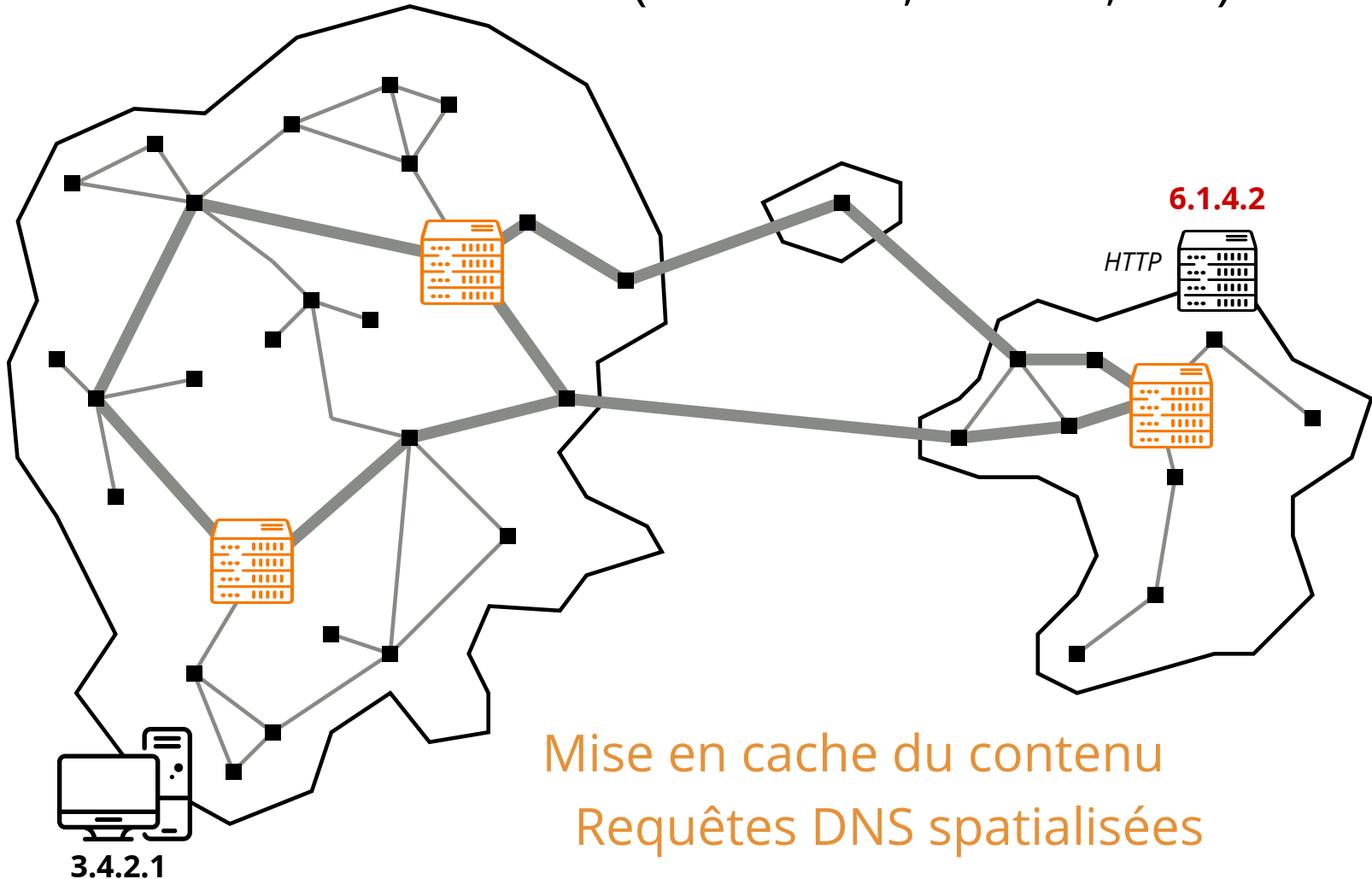
```
<!DOCTYPE html>
<html>

  <head>
    <meta charset="utf-8">
    <link rel="stylesheet" href="css/style.css">
    <script src="js/script.js"></script>
    <title>Test</title>
  </head>
  <body>
    <h1> Ma page web </h1>
    <p> Hello World </p>
  </body>
</html>
```

Protection anti-bot

Content Delivery Network (CDN)

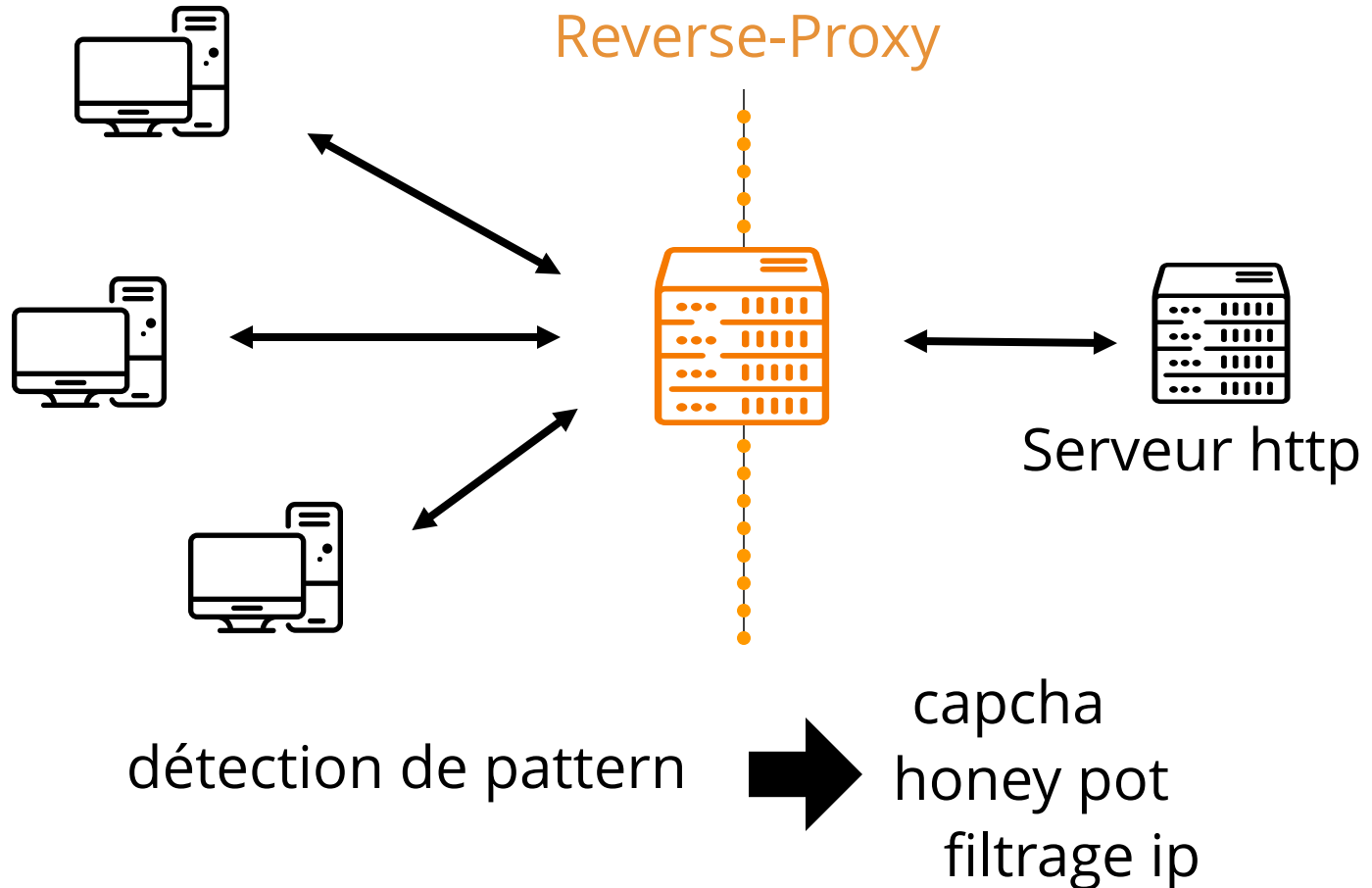
(CloudFlare, Akamai, etc.)



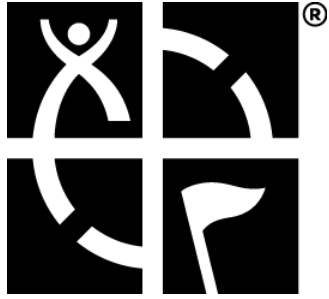
Protection anti-bot

Content Delivery Network (CDN)

(CloudFlare, Akamai, etc.)



GeoCaching



GRR Tenum (2015 - 2019) / Philippe Vidal

Récupérer **toutes les informations** sur les ~ **13000 géocaches** (1/page) de Normandie à un instant **t**

Site web : geocaching.com

Langage : Python

Framework : Scrapy

Type de récolte : Campagne

Format : json

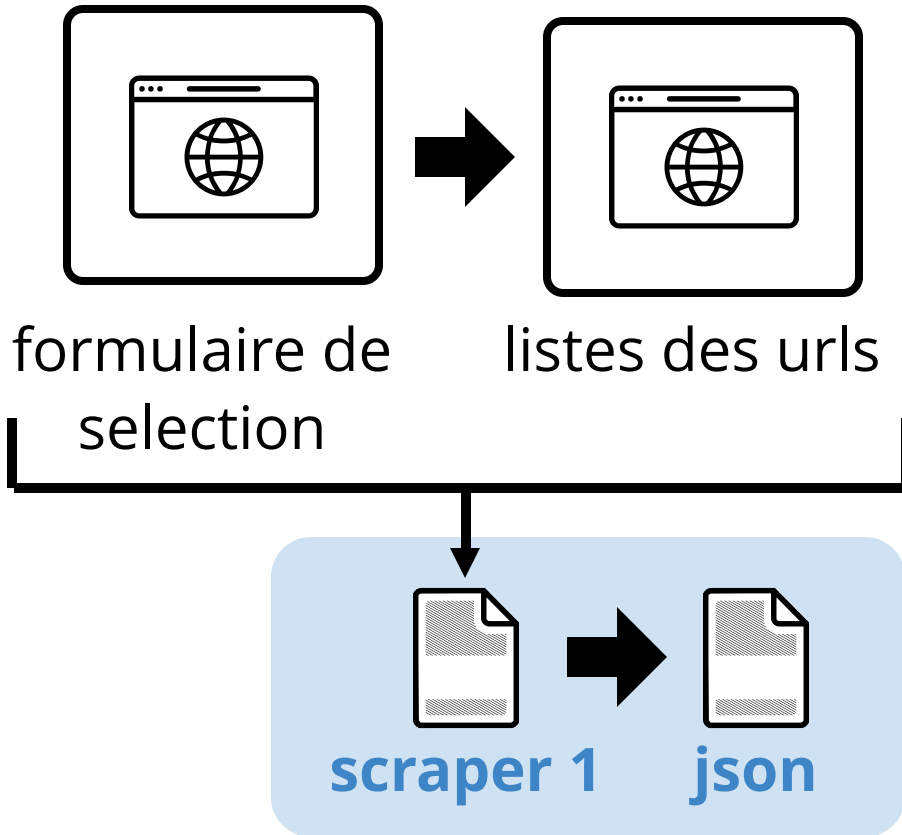
Difficultés :

- Pas d'API & limitation à 1000 résultats par recherche
- Mon premier scraping
- Login/Pwd pour accès à la donnée complète
- ASPX avec gestion des états précédents

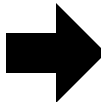


DOI [10.5281/zenodo.3542261](https://doi.org/10.5281/zenodo.3542261)

GeoCaching

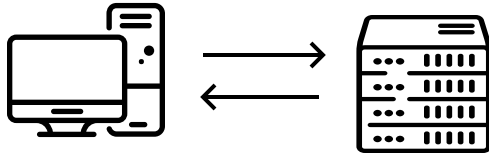


The screenshot shows a login page titled 'Connexion'. It features a blue button for 'Connexion avec Facebook'. Below this, there is a separator line with 'OU' in the center. Underneath, there are two input fields: 'E-MAIL OU NOM D'UTILISATEUR' with the placeholder text 'Entrez votre adresse e-mail ou votre nom d'utilis:' and 'MOT DE PASSE' with the placeholder text 'Veuillez saisir votre mot de passe'. A green 'Connexion' button is positioned below the input fields. At the bottom, there is a link: 'Vous avez oublié votre nom utilisateur ou votre mot de passe ?'.



GeoCaching

```
<form method="post" action="/nearest.aspx" onsubmit="javascript:return WebForm_OnSubmit();" id="aspnetForm">  
<div class="aspNetHidden">  
<input type="hidden" name="__EVENTTARGET" id="__EVENTTARGET" value="" />  
<input type="hidden" name="__EVENTARGUMENT" id="__EVENTARGUMENT" value="" />  
<input type="hidden" name="__LASTFOCUS" id="__LASTFOCUS" value="" />  
<input type="hidden" name="__VIEWSTATEFIELDSCOUNT" id="__VIEWSTATEFIELDSCOUNT" value="2" />  
<input type="hidden" name="__VIEWSTATE" id="__VIEWSTATE" value="wNwKV9Z" />  
<input type="hidden" name="__VIEWSTATE1" id="__VIEWSTATE1" value="jFMBVQ" />  
</div>
```



circulation token obligatoire

Recherche avancée

Les géocaches se trouvent dans le monde entier ! Rechercher des caches par code postal, émotif-clé, nom d'utilisateur ou code GC.

À vos débuts en Géocaching ? Prenez connaissance des [Premiers Pas](#) en Géocaching.

Rechercher :

Toutes les géocaches

Par :

Par Pays / État

Pays :

France

État/Province :

Normandie

Recherche de géocaches

navigation dans les pages & remplissage de formulaires

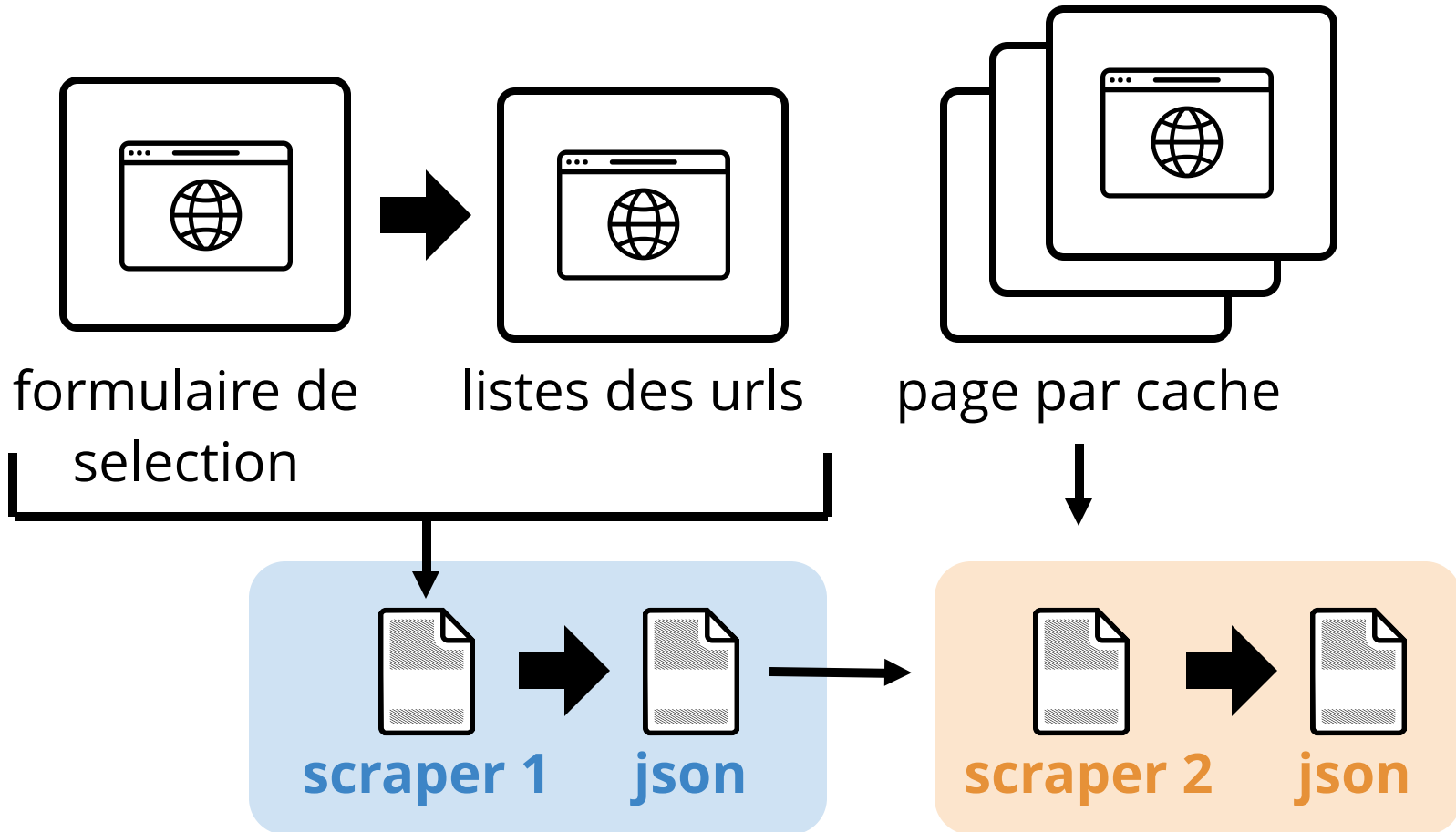
Total des Objets : 14513 - Page: 1 de 726 - <Précédent << <[1 2 3 4 5 6 7 8 9 10]> >> Suivant >

			Description	Icônes	(D / T)	Placée	Dernière découverte
<input type="checkbox"/>				[CIC5] CITO : protégeons la nature !!! par Team Christmas GC8DW5B Normandie, France		1/2 12/22/2019	NOUVEAU !
<input type="checkbox"/>				Christmas is Coming #5 Last Christmas par call-amar GC8E0TN Normandie, France		1/1 12/21/2019	NOUVEAU !
<input type="checkbox"/>				Le Noël des Géocacheurs Normands - 6ème Edition par Coulanpic GNAC GC8FN2V Normandie,		1/1 12/15/2019	NOUVEAU !



URL des 726 * 20 pages

GeoCaching



GeoCaching

N commentaires par cache ...

161 visites enregistrées

😄 147 😞 1 🗺️ 1 📍 1 🗝️ 1

[Voir le journal](#) | [Regarder la galerie d'images de 88 images](#) Attention ! Des spoilers (indice photo) peuvent faire partie des descriptions ou des liens.

Leloulou
Premium Member

😄 7179

Found it 11/2/2019

Week end en amoureux avec ma femme adorée Tem77.
Au programme du week end...pluie et tempête.
Bon on décide de remonter la Seine de Paris à Rouen, puis de revenir à la maison en faisant les caches en drive in ou pas très loin de la voiture. 2° objectif: continuer de verdir les tranches du challenge 360° (Gc4RNY3).
On aura galéré un peu pour faire verdir le cheker mais cela nous aura appris vraiment pas mal de choses. Sur place nous ferons la deuxième étape et la réponse envoyé à l'owner.
Nous finirons notre soirée au resto pour retourner nous coucher à notre hôtel.
Merci pour la cache, et la découverte de toutes ces spécialités.

 [Tem à Rouen à déguster du local](#)

[Afficher le log](#)

Tem77
Premium Member

😄 3076

Found it 11/2/2019

Un long week end devant nous ... pluie annoncée sur toute la France...mais pas envie de rester à la maison...opération recherche de boites pour valider quelques portions de camembert, organisée de main de maître par mon Loulou et mari d amour♥
C est parti ... bottes et kway dans la voiture et direction la Seine et la manche🇫🇷
2eme nuit ... stop sur Rouen
Merci pour la virtuelle et pour la découverte de le endroit

[Afficher le log](#)

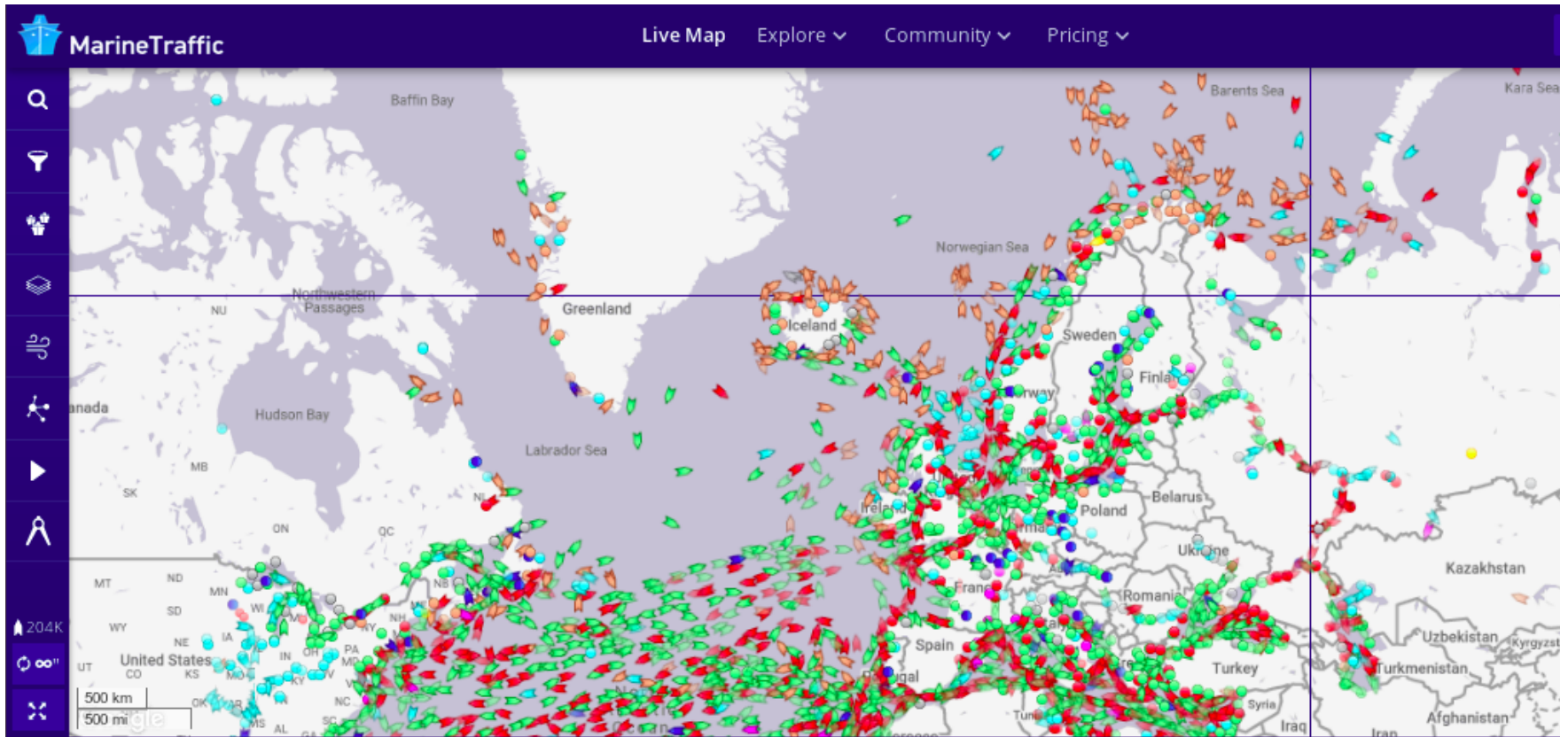


Analyse textuelle (work in progress)

Arctique & système monde



Intégration des zones arctiques dans le système monde
Flux maritime : Marine Traffic



Trafic maritime

Liste des bateaux dans un port (Murmansk) :

Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VYAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VYAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

SHIP_ID: "351170"
SHIPNAME: "GULDRANGUR"
PORT_ID: "2244"
PORT_NAME: "MURMANSK"
CENTERX: "33.061"
CENTERY: "68.98412"
COUNTRY_CODE: "RU"
IMO: "8315944"
TIMESTAMP_UTC: 1573767600
FROM_NOANCH_ID: null
FROM_NOANCH_NAME: null
TIMEZONE: "3"
MOVE_TYPE_NAME: "DEPARTURE"
MMSI: "273554600"

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VTAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

SHIP_ID: "351170"
SHIPNAME: "GULDRANGUR"
PORT_ID: "2244"
PORT_NAME: "MURMANSK"
CENTERX: "33.061"
CENTERY: "68.98412"
COUNTRY_CODE: "RU"
IMO: "8315944"
TIMESTAMP_UTC: 1573767600
FROM_NOANCH_ID: null
FROM_NOANCH_NAME: null
TIMEZONE: "3"
MOVE_TYPE_NAME: "DEPARTURE"
MMSI: "273554600"

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VTAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

SHIP_ID: "351170"
SHIPNAME: "GULDRANGUR"
PORT_ID: "2244"
PORT_NAME: "MURMANSK"
CENTERX: "33.061"
CENTERY: "68.98412"
COUNTRY_CODE: "RU"
IMO: "8315944"
TIMESTAMP_UTC: 1573767600
FROM_NOANCH_ID: null
FROM_NOANCH_NAME: null
TIMEZONE: "3"
MOVE_TYPE_NAME: "DEPARTURE"
MMSI: "273554600"

Dernières positions d'un bateau (Guldrangur) :

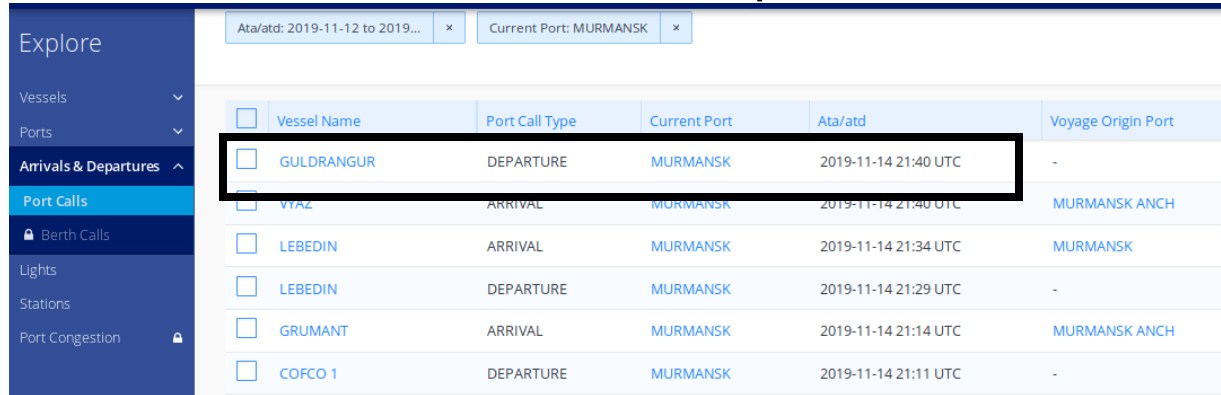
MarineTraffic Live Map Explore Community Pricing Vessel, Port, etc.

Quick Search: GULDRANGUR x Add Filter

<input type="checkbox"/>	Vessel Name	Event	Timestamp	Latitude	Longitude
<input type="checkbox"/>	GULDRANGUR	Departure	2019-11-14 21:40 UTC	-	-
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-14 21:09 UTC	68.95415	33.0191
<input type="checkbox"/>	GULDRANGUR	Noon position	2019-11-14 09:03 UTC	68.94379	33.00941
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-13 21:11 UTC	68.94376	33.00934

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

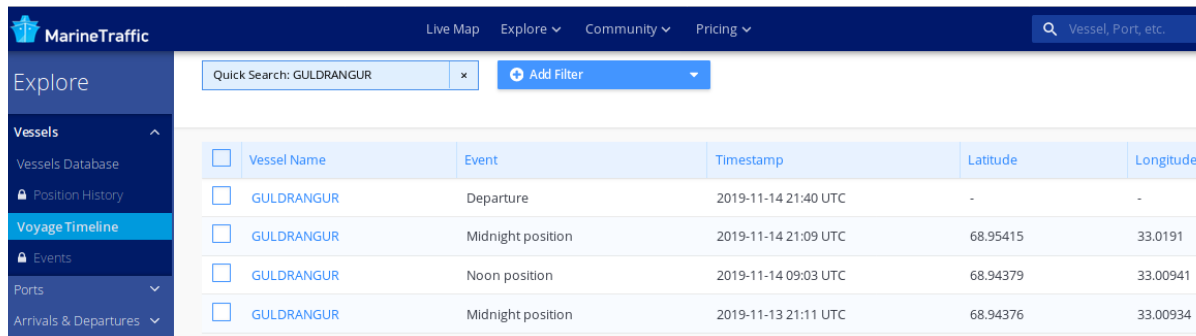


Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VTAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

```
SHIP_ID: "351170"  
SHIPNAME: "GULDRANGUR"  
PORT_ID: "2244"  
PORT_NAME: "MURMANSK"  
CENTERX: "33.061"  
CENTERY: "68.98412"  
COUNTRY_CODE: "RU"  
IMO: "8315944"  
TIMESTAMP.UTC: 1573767600  
FROM_NOANCH_ID: null  
FROM_NOANCH_NAME: null  
TIMEZONE: "3"  
MOVE_TYPE_NAME: "DEPARTURE"  
MMSI: "273554600"
```

Dernières positions d'un bateau (Guldrangur) :



MarineTraffic Live Map Explore Community Pricing Vessel, Port, etc.

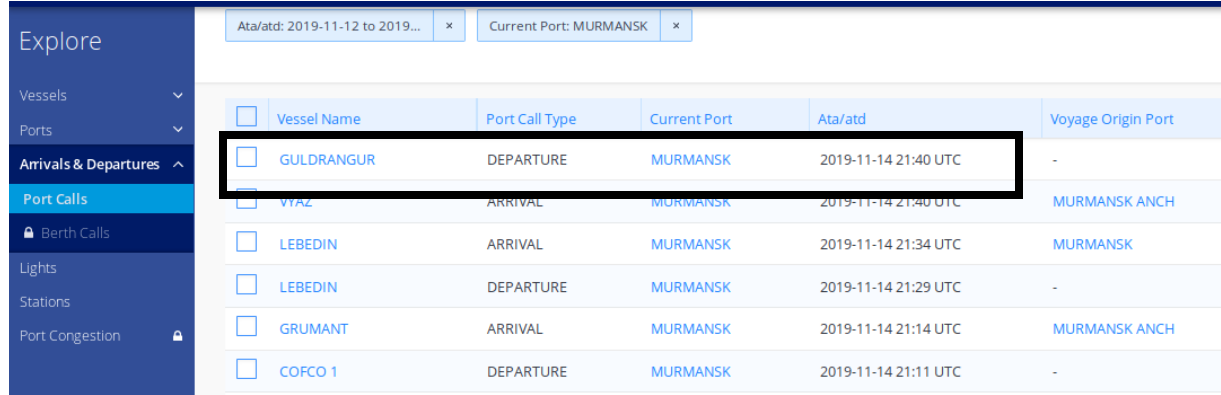
Quick Search: GULDRANGUR x Add Filter

<input type="checkbox"/>	Vessel Name	Event	Timestamp	Latitude	Longitude
<input type="checkbox"/>	GULDRANGUR	Departure	2019-11-14 21:40 UTC	-	-
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-14 21:09 UTC	68.95415	33.0191
<input type="checkbox"/>	GULDRANGUR	Noon position	2019-11-14 09:03 UTC	68.94379	33.00941
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-13 21:11 UTC	68.94376	33.00934

```
SHIP_ID: "351170"  
MMSI: "273554600"  
EVENT_ID: "1"  
TIMESTAMP: 1573765740  
SHIPNAME: "GULDRANGUR"  
EVENT_NAME: "Midnight position"  
LAT: "68.95415"  
LON: "33.0191"  
OLD_TIMESTAMP: "2019-11-14 21:09:00"  
TIMEZONE: "3"
```

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

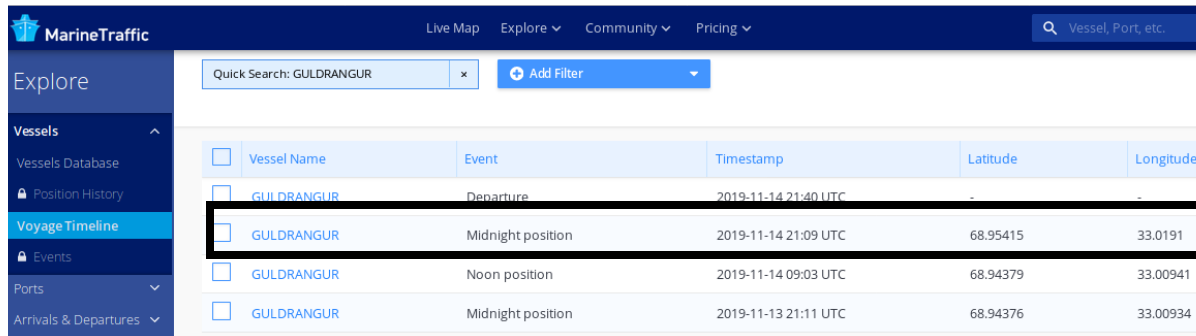


Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VTAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

```
SHIP_ID: "351170"  
SHIPNAME: "GULDRANGUR"  
PORT_ID: "2244"  
PORT_NAME: "MURMANSK"  
CENTERX: "33.061"  
CENTERY: "68.98412"  
COUNTRY_CODE: "RU"  
IMO: "8315944"  
TIMESTAMP.UTC: 1573767600  
FROM_NOANCH_ID: null  
FROM_NOANCH_NAME: null  
TIMEZONE: "3"  
MOVE_TYPE_NAME: "DEPARTURE"  
MMSI: "273554600"
```

Dernières positions d'un bateau (Guldrangur) :



MarineTraffic Live Map Explore Community Pricing Vessel, Port, etc.

Quick Search: GULDRANGUR x Add Filter

<input type="checkbox"/>	Vessel Name	Event	Timestamp	Latitude	Longitude
<input type="checkbox"/>	GULDRANGUR	Departure	2019-11-14 21:40 UTC	-	-
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-14 21:09 UTC	68.95415	33.0191
<input type="checkbox"/>	GULDRANGUR	Noon position	2019-11-14 09:03 UTC	68.94379	33.00941
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-13 21:11 UTC	68.94376	33.00934

```
SHIP_ID: "351170"  
MMSI: "273554600"  
EVENT_ID: "1"  
TIMESTAMP: 1573765740  
SHIPNAME: "GULDRANGUR"  
EVENT_NAME: "Midnight position"  
LAT: "68.95415"  
LON: "33.0191"  
OLD_TIMESTAMP: "2019-11-14 21:09:00"  
TIMEZONE: "3"
```

Trafic maritime

Liste des bateaux dans un port (Murmansk) :

Ata/atd: 2019-11-12 to 2019... x Current Port: MURMANSK x

<input type="checkbox"/>	Vessel Name	Port Call Type	Current Port	Ata/atd	Voyage Origin Port
<input type="checkbox"/>	GULDRANGUR	DEPARTURE	MURMANSK	2019-11-14 21:40 UTC	-
<input type="checkbox"/>	VTAZ	ARRIVAL	MURMANSK	2019-11-14 21:40 UTC	MURMANSK ANCH
<input type="checkbox"/>	LEBEDIN	ARRIVAL	MURMANSK	2019-11-14 21:34 UTC	MURMANSK
<input type="checkbox"/>	LEBEDIN	DEPARTURE	MURMANSK	2019-11-14 21:29 UTC	-
<input type="checkbox"/>	GRUMANT	ARRIVAL	MURMANSK	2019-11-14 21:14 UTC	MURMANSK ANCH
<input type="checkbox"/>	COFCO 1	DEPARTURE	MURMANSK	2019-11-14 21:11 UTC	-

```
SHIP_ID: "351170"  
SHIPNAME: "GULDRANGUR"  
PORT_ID: "2244"  
PORT_NAME: "MURMANSK"  
CENTERX: "33.061"  
CENTERY: "68.98412"  
COUNTRY_CODE: "RU"  
IMO: "8315944"  
TIMESTAMP.UTC: 1573767600  
FROM_NOANCH_ID: null  
FROM_NOANCH_NAME: null  
TIMEZONE: "3"  
MOVE_TYPE_NAME: "DEPARTURE"  
MMSI: "273554600"
```

Dernières positions d'un bateau (Guldrangur) :

Quick Search: GULDRANGUR x Add Filter

<input type="checkbox"/>	Vessel Name	Event	Timestamp	Latitude	Longitude
<input type="checkbox"/>	GULDRANGUR	Departure	2019-11-14 21:40 UTC	-	-
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-14 21:09 UTC	68.95415	33.0191
<input type="checkbox"/>	GULDRANGUR	Noon position	2019-11-14 09:03 UTC	68.94379	33.00941
<input type="checkbox"/>	GULDRANGUR	Midnight position	2019-11-13 21:11 UTC	68.94376	33.00934

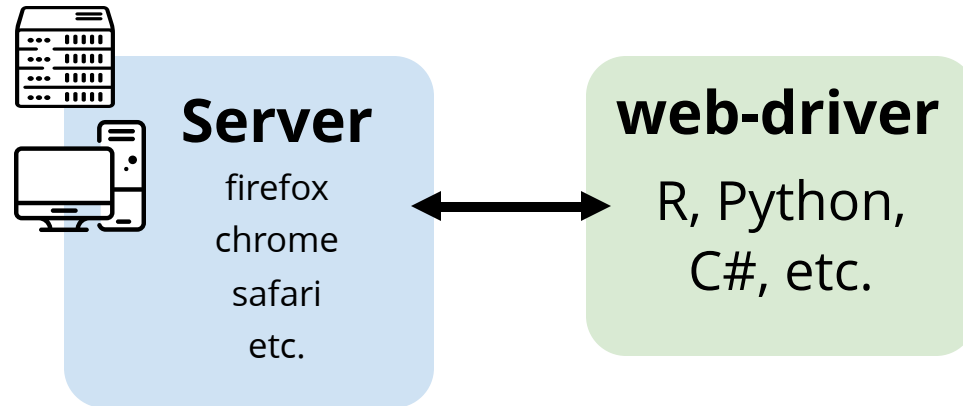
```
SHIP_ID: "351170"  
MMSI: "273554600"  
EVENT_ID: "1"  
TIMESTAMP: 1573765740  
SHIPNAME: "GULDRANGUR"  
EVENT_NAME: "Midnight position"  
LAT: "68.95415"  
LON: "33.0191"  
OLD_TIMESTAMP: "2019-11-14 21:09:00"  
TIMEZONE: "3"
```

Même approche que pour Flightradar ?

403 Forbidden

Mr Robot

Selenium tools

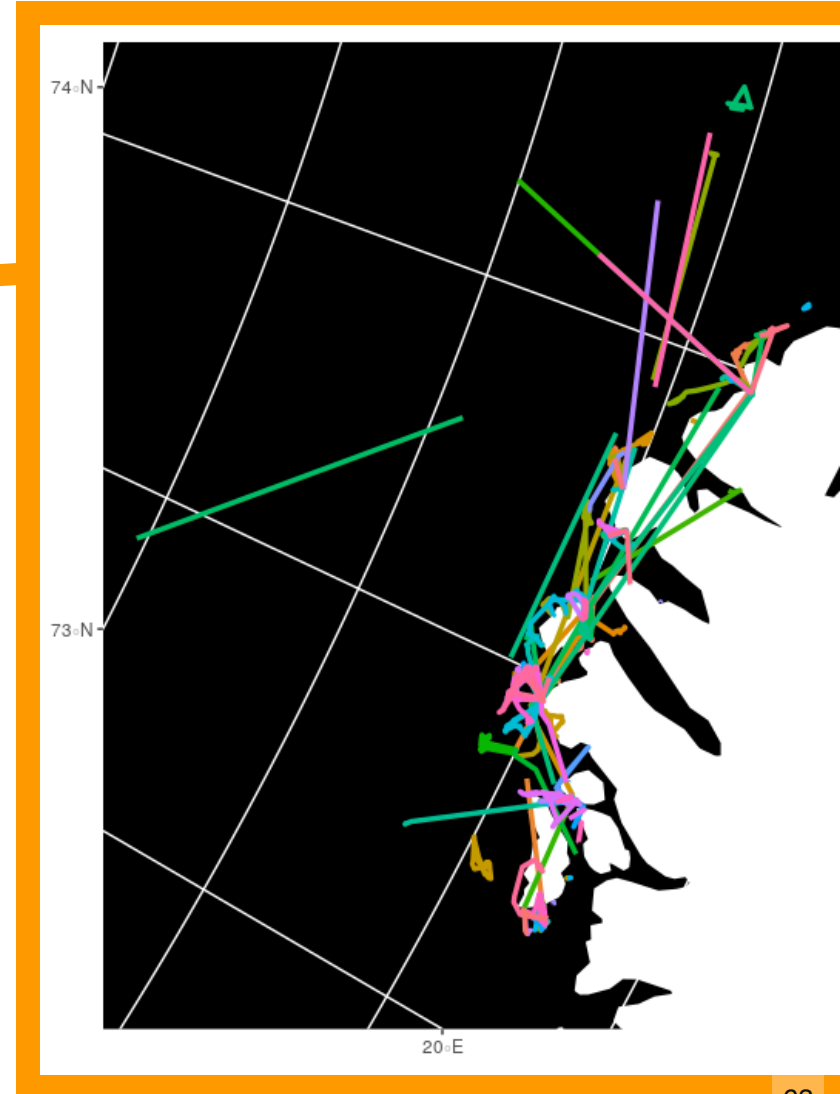
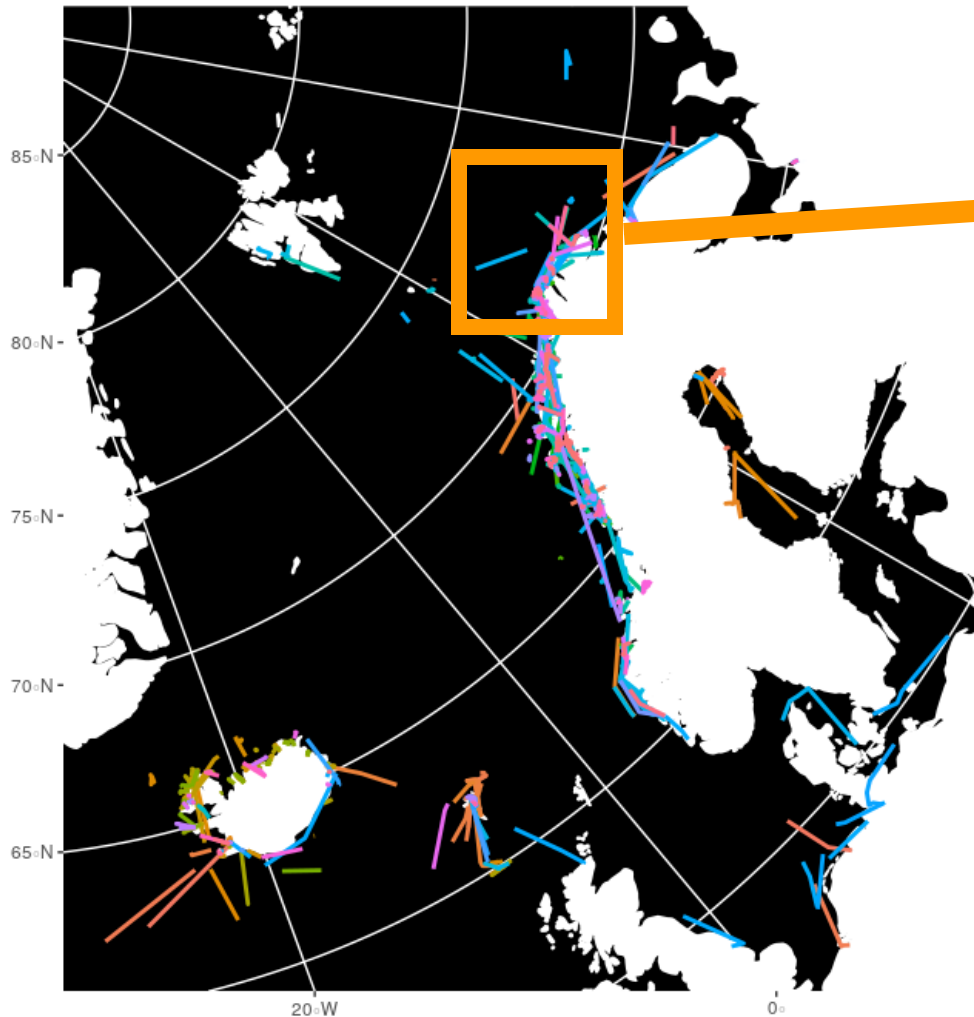


API générique, multi-langage (Java, Python, R, etc.) pour piloter (envoyer des commandes) à un navigateur sans utiliser l'interface graphique.

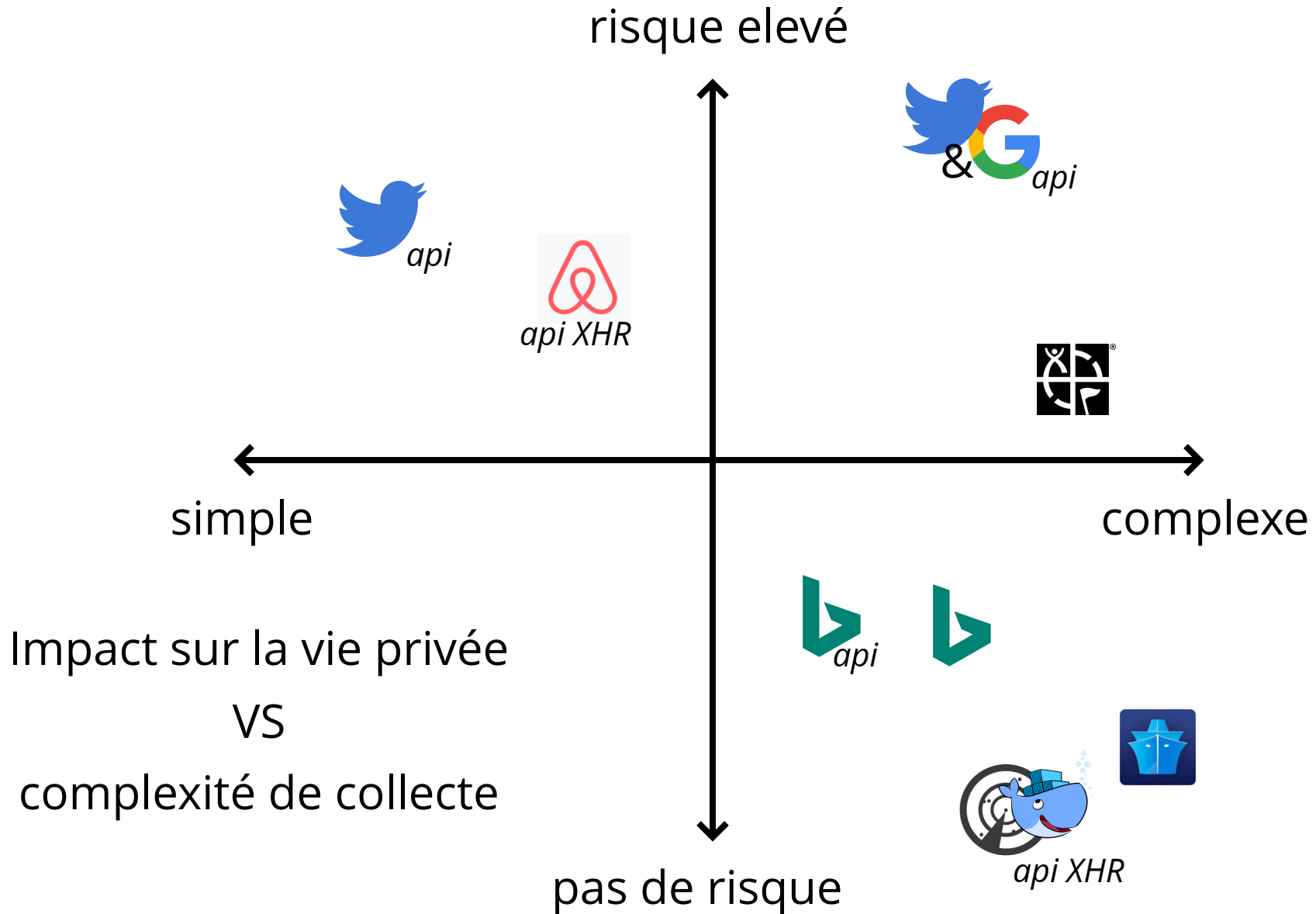
Tutorial avec R & Docker

Marine Traffic

Trajectoire des bateaux sur 2 jours :



Ethique & Technique



Un tryptique asymétrique

(Schéma simpliste)

Utilise ou contribue au service ==> participe à la création de données (+/- consciente)

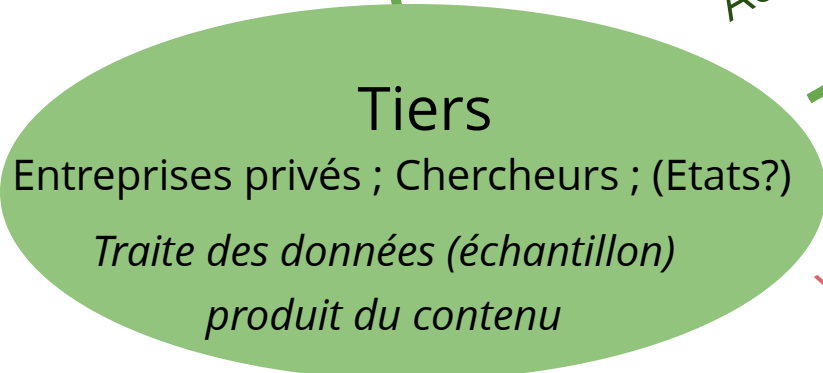


Propose des services & stock les données
(publicité)



Demande rarement
le consentement éclairé

Accède aux données consultable
sur le site (webscrapping)



Vend ou donne des échantillons
contrôle l'utilisation (API)

Questions ?

Merci de votre attention !

<https://slides.com/sebastienreycoyrehourcq/cistscrapping>



The screenshot shows the top of the Le Monde website. The masthead features the 'Le Monde' logo in a large, black, gothic-style font. To the left is a small thumbnail of the newspaper and a link to 'Consulter le journal'. Below the masthead is a horizontal navigation bar with categories: ACTUALITÉS, ÉCONOMIE, VIDÉOS, OPINIONS, CULTURE, M LE MAG, and SE. The main content area has a sub-header 'PIXELS • VIE PRIVÉE' and a 'Partage' link. The headline reads 'La surveillance des réseaux sociaux contre la fraude fiscale adoptée à l'Assemblée'. A sub-headline states: 'Les députés ont donné leur feu vert pour que les services fiscaux et douaniers aspirent massivement les données des Français afin de détecter les fraudeurs.' At the bottom, it says 'Par Morgane Tual · Publié hier à 23h03, mis à jour à 11h23'.

le 13/11/2019



The screenshot shows the top of The Wall Street Journal website. The masthead features 'THE WALL STREET JOURNAL.' in a large, black, serif font. To the right are links for 'Subscribe' and 'Sign In'. Below the masthead is a horizontal navigation bar with categories: Home, World, U.S., Politics, Economy, Business, Tech, Markets, Opinion, Life & Arts, Real Estate, and WSJ. Magazine. A search icon is on the far right. The main content area has a sub-header '◆ WSJ NEWS EXCLUSIVE | TECH' and a headline: 'Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans'. A sub-headline states: 'Search giant is amassing health records from Ascension facilities in 21 states; patients not yet informed'.

