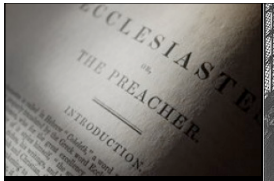
Big Data – Bad Data

Rayna Stamboliyska, PhD

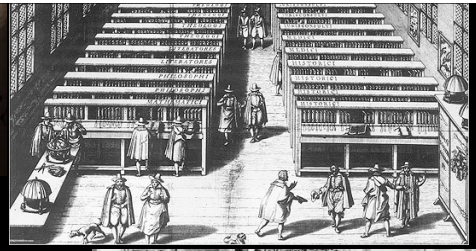
Paris Descartes University
RS Strategy

XXXL
Data

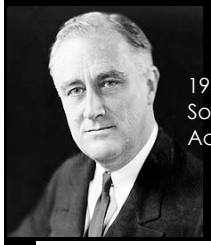
XXXL
Issues



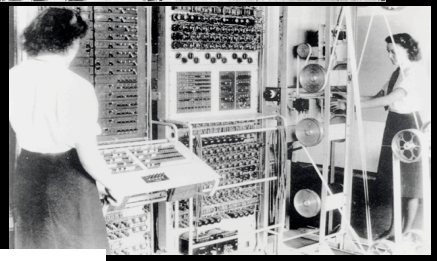
300 years BC: "of making many books there is no end"



"Is there anywhere on earth exempt from these swarms of new books?"



1935-1937, Social Security Act



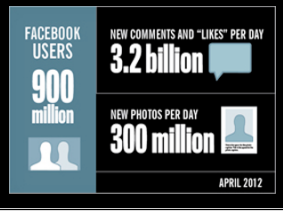
1943-1945, 'Colossus' & Enigma

1 Introduction

Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of *big data*. When data sets do not fit in main memory (*in*

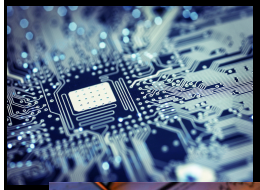
1997: first mention of 'big data' in *Application-Controlled Demand Paging for Out-of-Core Visualization* (NASA)

...



No, Big Data isn't new... but

What is it?

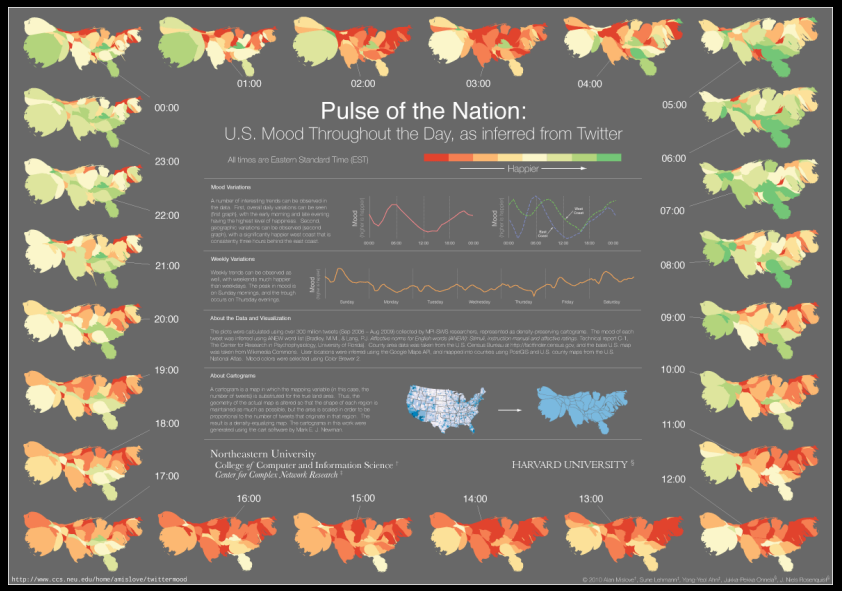


Maximizing computation power and algorithmic accuracy

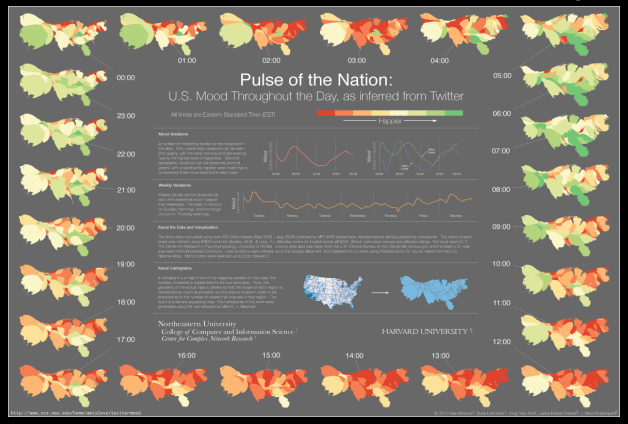
Drawing on a range of tools to analyse and compare large datasets

Believing large datasets give greater objectivity and accuracy

No, Data doesn't speak of itself

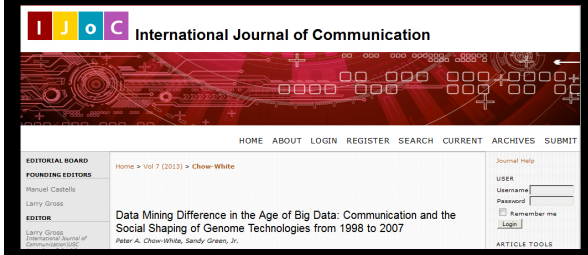


Numbers never lie... o rly?

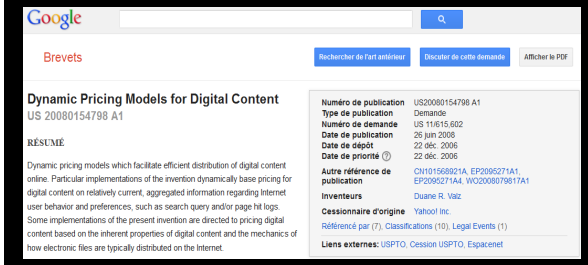


* 51% of web traffic is non-human
* there are 30 million fake Twitter accounts

Yes, Big Data discriminates between social groups

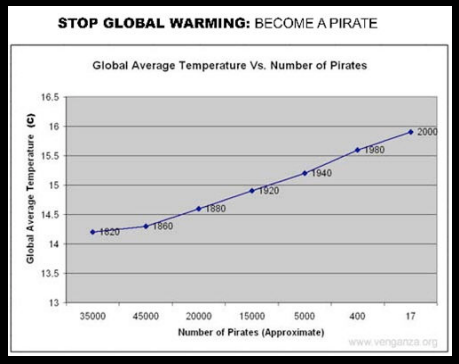


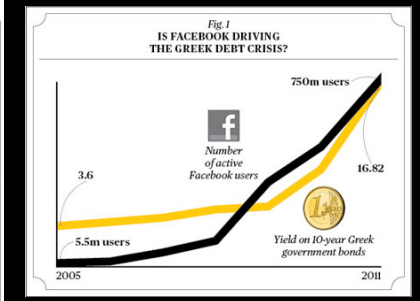
Scientists are allowing their assumptions about race to shape their big-data genomics research



If your past buying history indicates you are more likely to pay top euro for trousers, your starting price the next time you shop for fooclothes online might be considerably higher.

Big Data Fundamentalism?





There's a word for that: **apophenia**!

… Does big data change the erroneous 'correlation = causation' dynamic?

Math majors, rejoice!

"With big data techniques, you can get much closer to being able to predict causal relations."

I beg to disagree

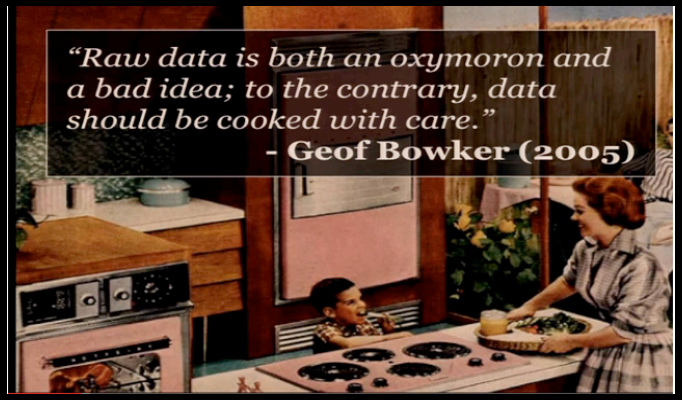
Your question and your data are not random entities.

Big Data needs several steps:

- of preparation

And...

If data are somehow subject to us, we are also subject to data



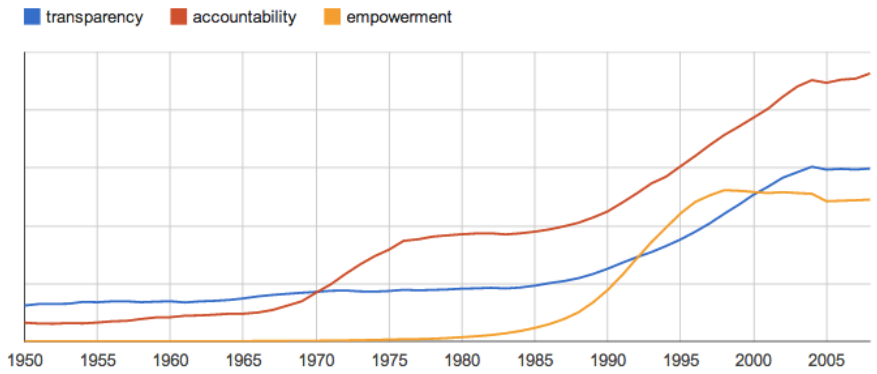
Open Data
as sovereignty.

Romain Lacombe
Etalab (data.gouv.fr)

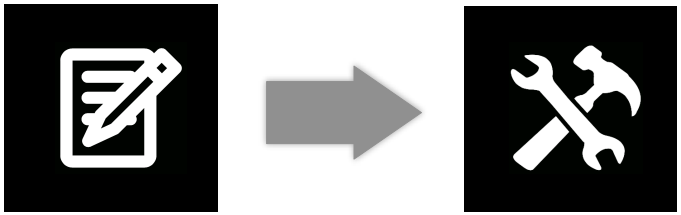
Code is **law**
Architecture is
politics

Data is
democracy

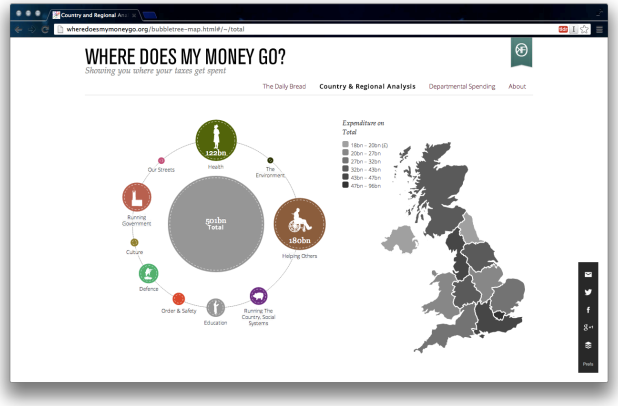
Answer to rising **aspirations**



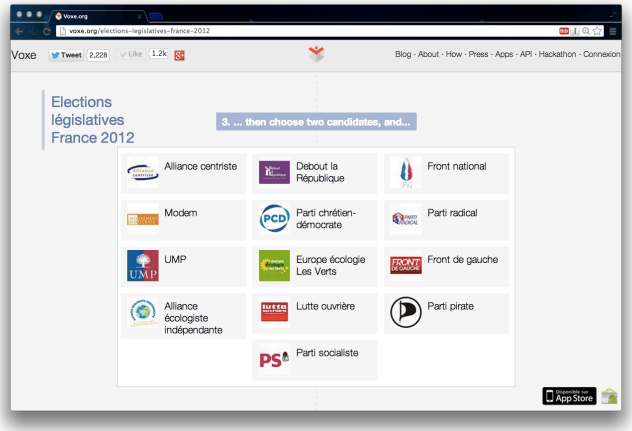
Tool for **empowerment**



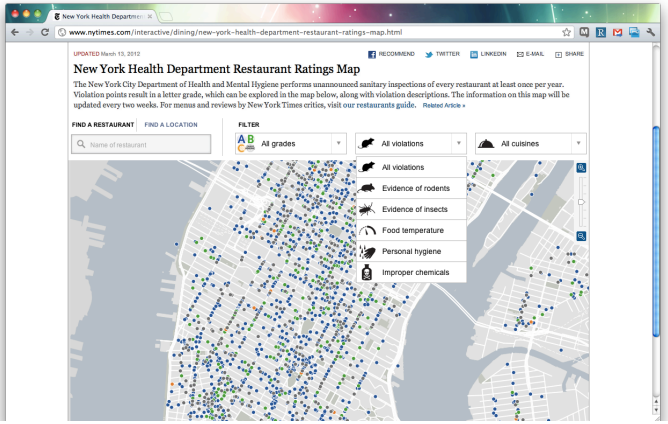
Perspective
on **collective action**



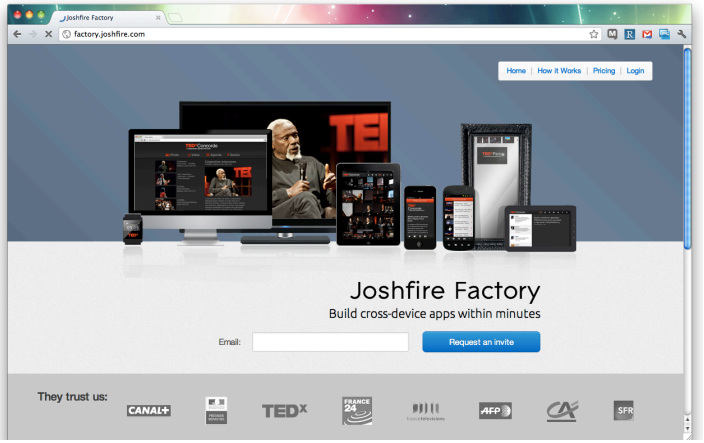
Informing
individual choices



Transparency as **regulation**

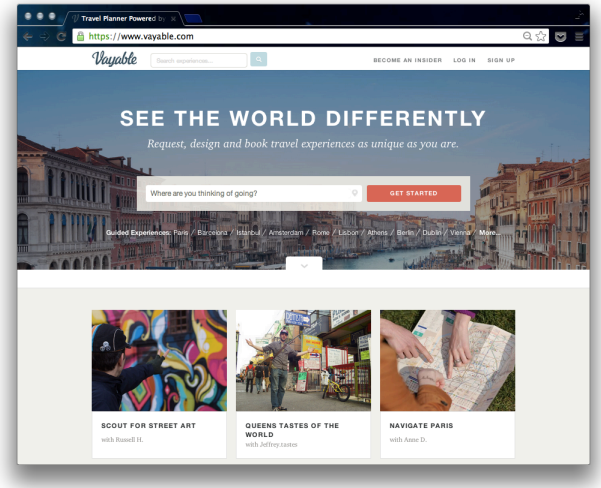


Innovation as **reinvention**




Sharing
economy
SOYEZ LIBRE
COMME L'AIR
BT 349 TF

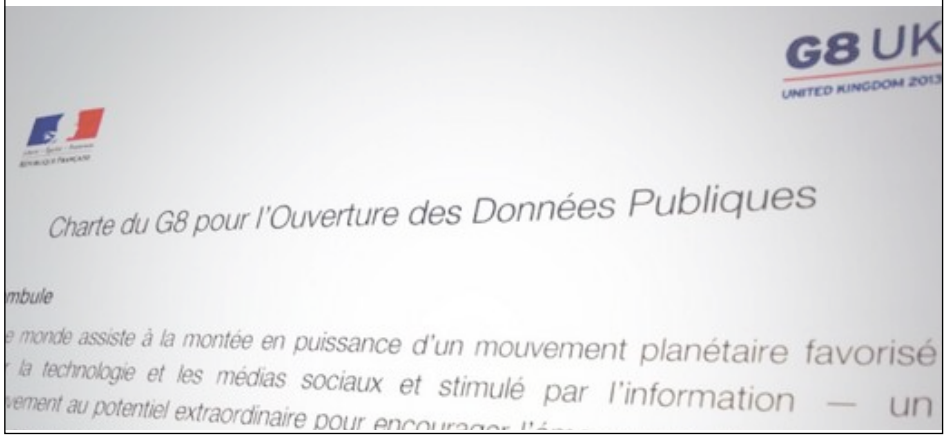
Shift to **experience**



Connected
cities and policies

Governing
data?

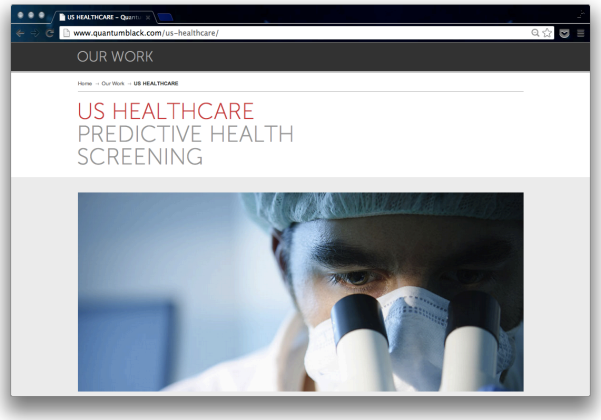
Open **government data**



From repository to
open platform

Openness drives **innovation**

Innovation is **sovereignty**



Code is **law**
Architecture is **politics**
Data is **democracy**
Openness is
sovereignty

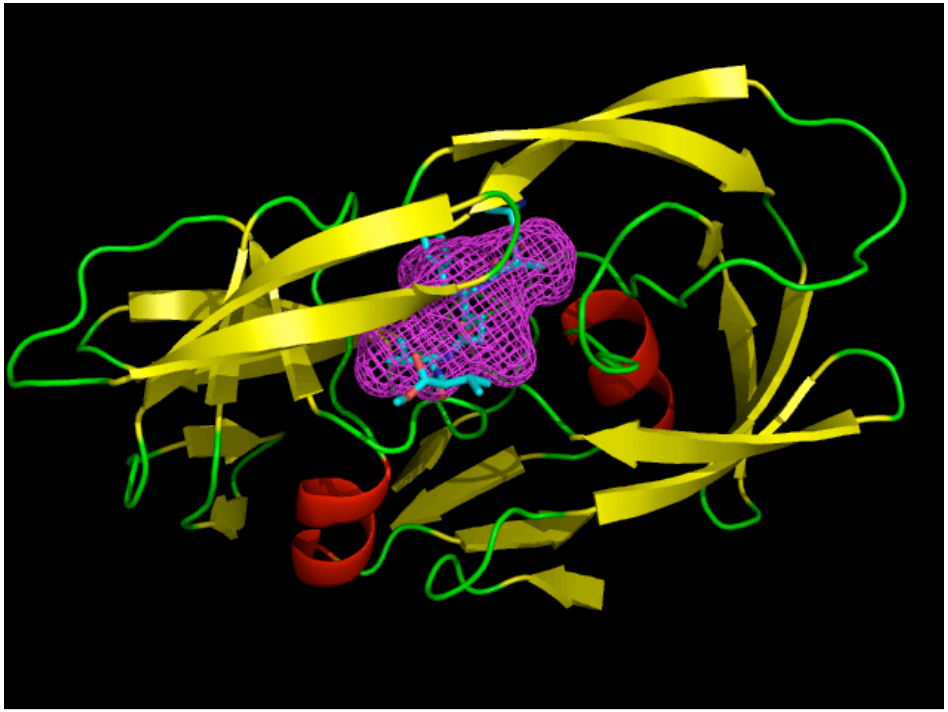
Thank you.

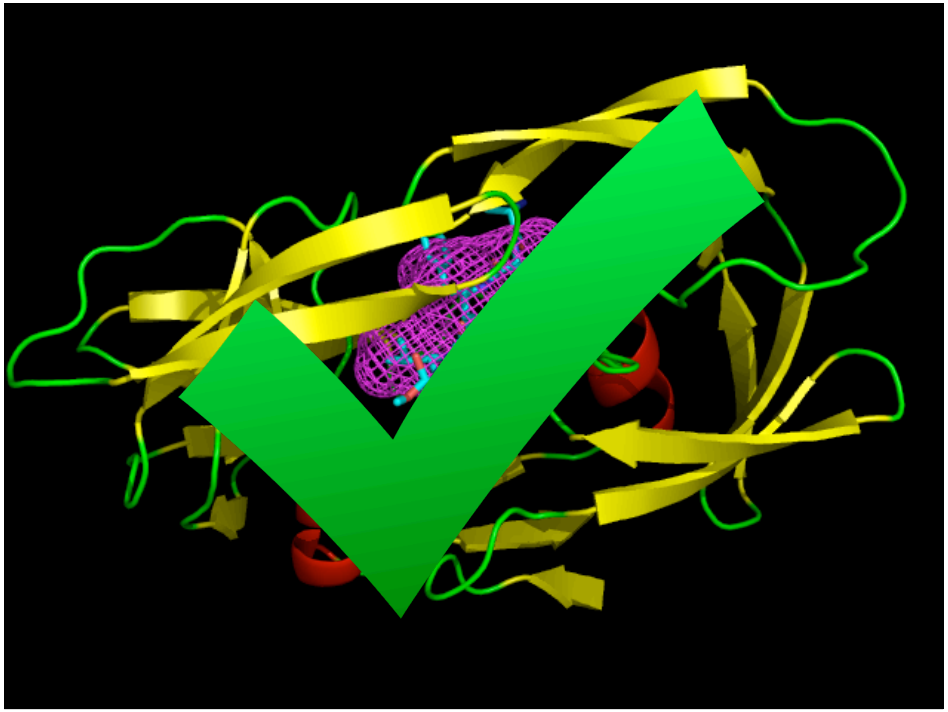
Romain Lacombe
Etalab (data.gouv.fr)

:snips

Rand Hindi, PhD
@randhindi


BIG OR BAD?

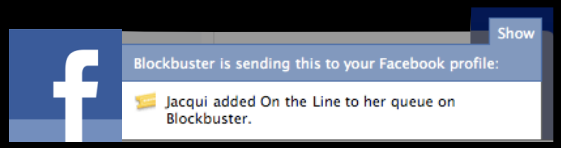





PRISM


PRISI

BEACON


Show
Blockbuster is sending this to your Facebook profile:
Jacqui added On the Line to her queue on Blockbuster.

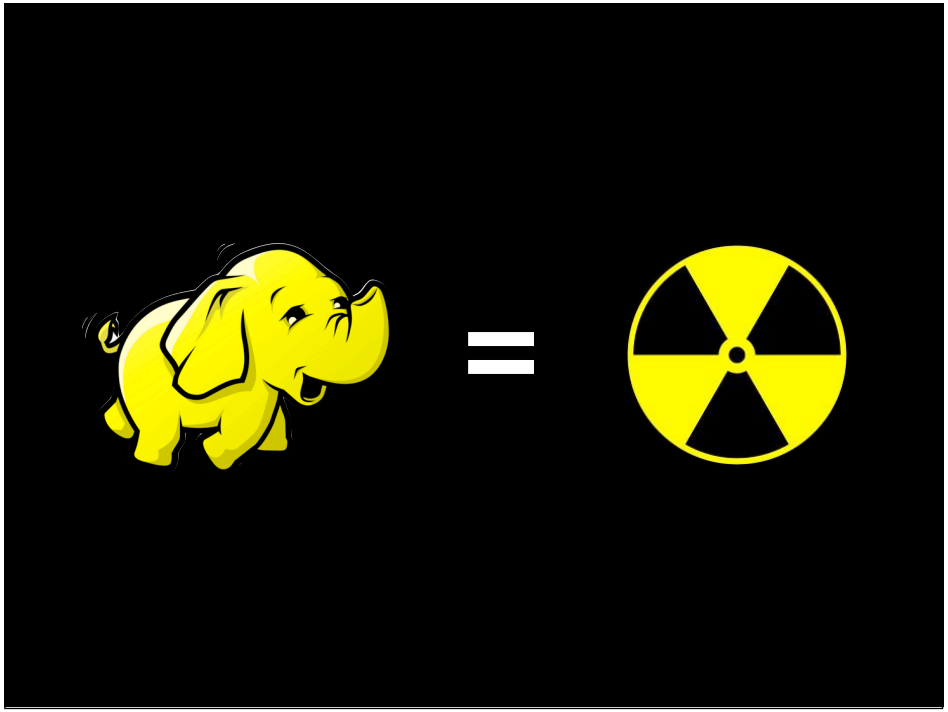

BEACON
Show
Blockbuster is sendi
our Facebook profile:
Jacqui added On the
queue on
Blockbuster.


Carrier
3:39 PM
TRANQUILIEN
DÉPART
Paris Austerlitz
ARRIVÉE
Gare arrivée
DATE
Maintenant
16:57
GATA
vers Montigny Beauchamp
16:59
BALI
vers Bretigny
17:01
ELAO
vers Saint-Martin d'Etampes
17:02
PAUL
vers Pont du Garigliano Hôp...
17:03
BALI
vers Bretigny


TRANQUILIEN
Paris Austerlitz
ARRIVÉE
DATE
Maintenant
16:57
vers Montigny Beauchamp
16:59
vers Bretigny
17:01
vers Saint-Martin d'Etampes
17:02
vers Pont du Garigliano Hôp...
17:03
vers Bretigny


TARGET


TARGE



1. AGGREGATE

1. AGGREGATE

2. ANONYMIZE

1. AGGREGATE
2. ANONYMIZE
3. OPT-OUT / IN

WOULD YOU LIKE IT
IF SOMEONE ELSE
DID THAT TO YOU ?


HAVE
FUN AND
DON’T
FUCK UP