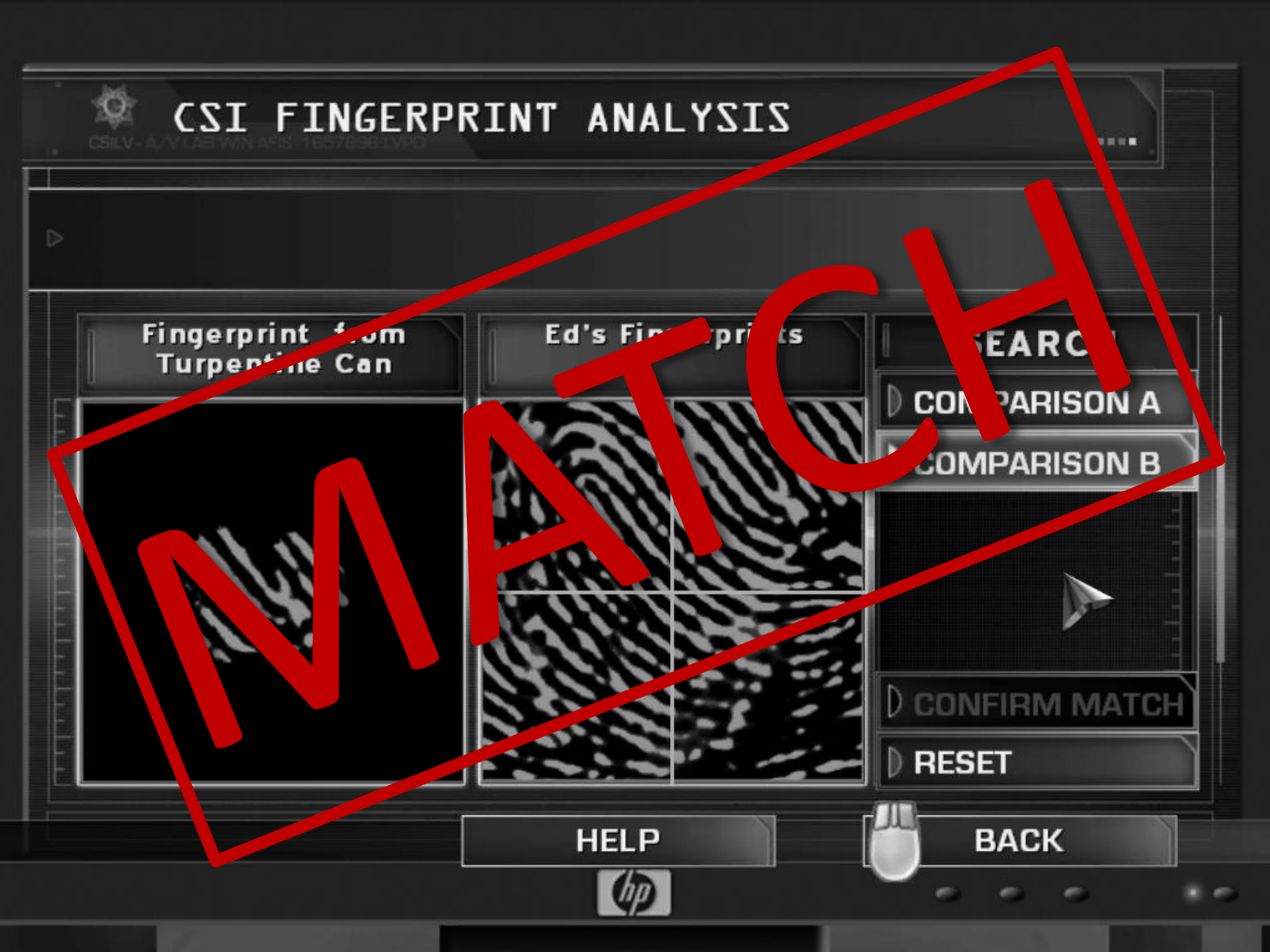
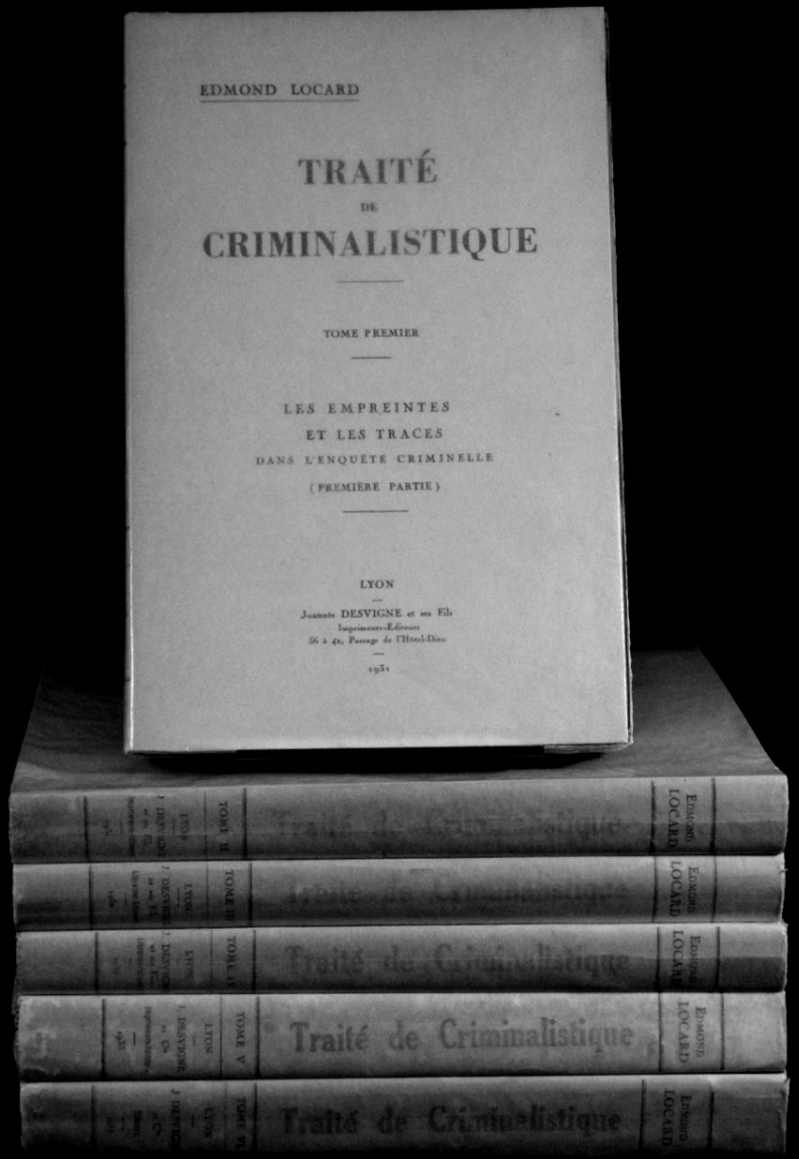
Big but personal data

How our behavior makes us unique

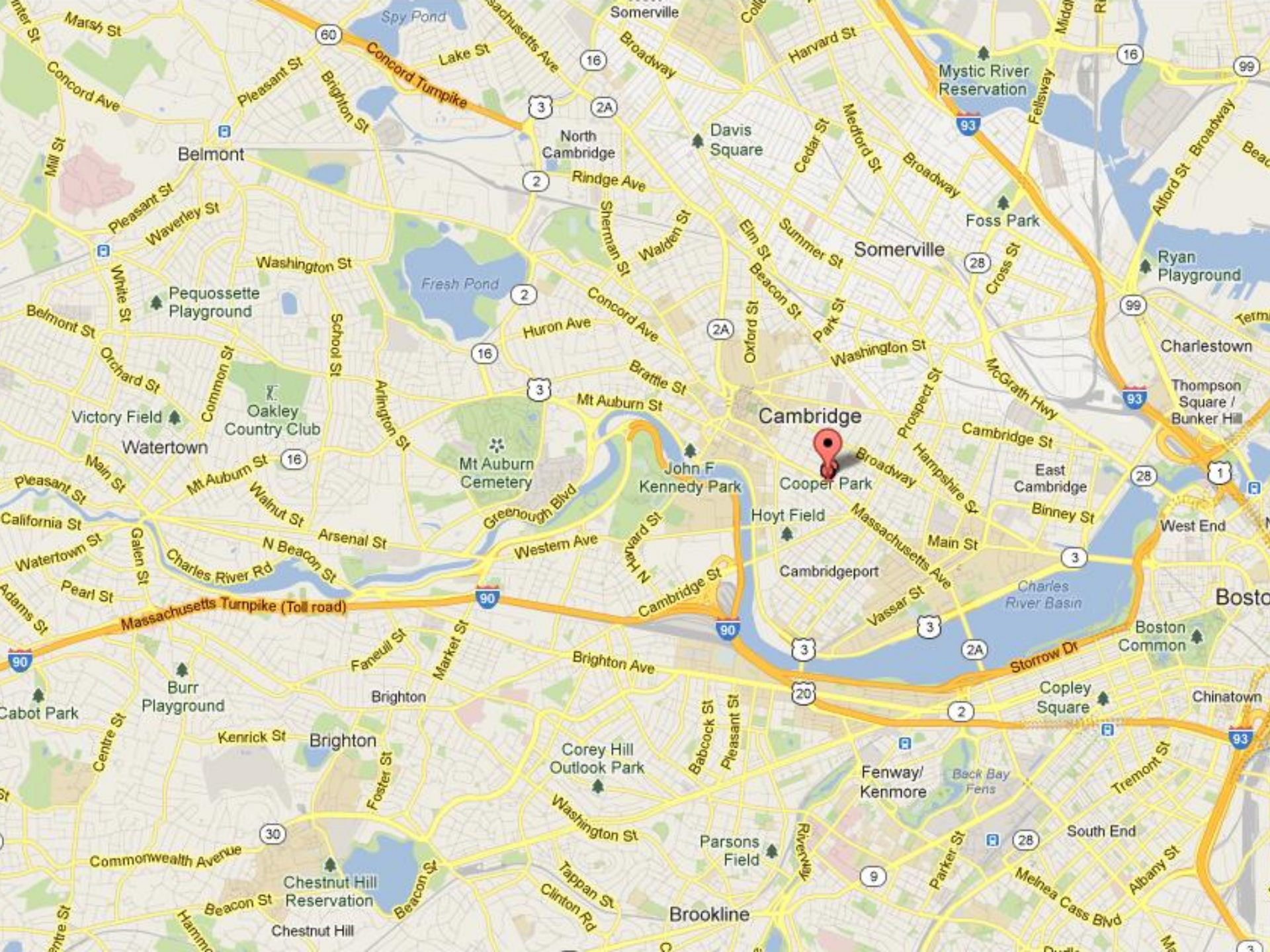
Yves-Alexandre de Montjoye
MIT Media Lab – Human Dynamics

CSI:
CRIME SCENE INVESTIGATION
Cleveland's
WB
wbnx-tv

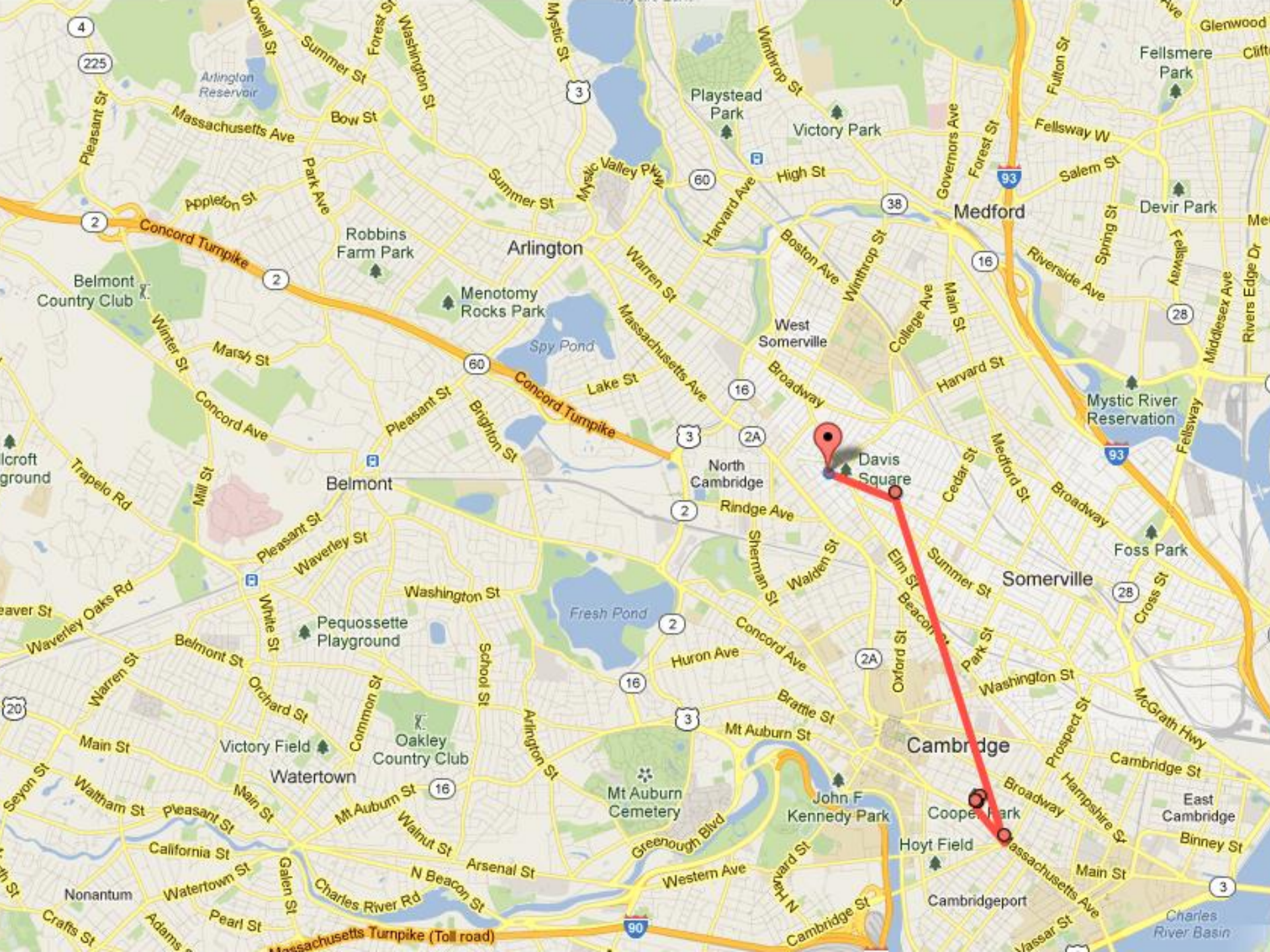

CSI FINGERPRINT ANALYSIS
Fingerprint from Turpentine Can
Ed's Fingerprints
SEARCH
COMPARISON A
COMPARISON B
CONFIRM MATCH
RESET
HELP
BACK
MATCH

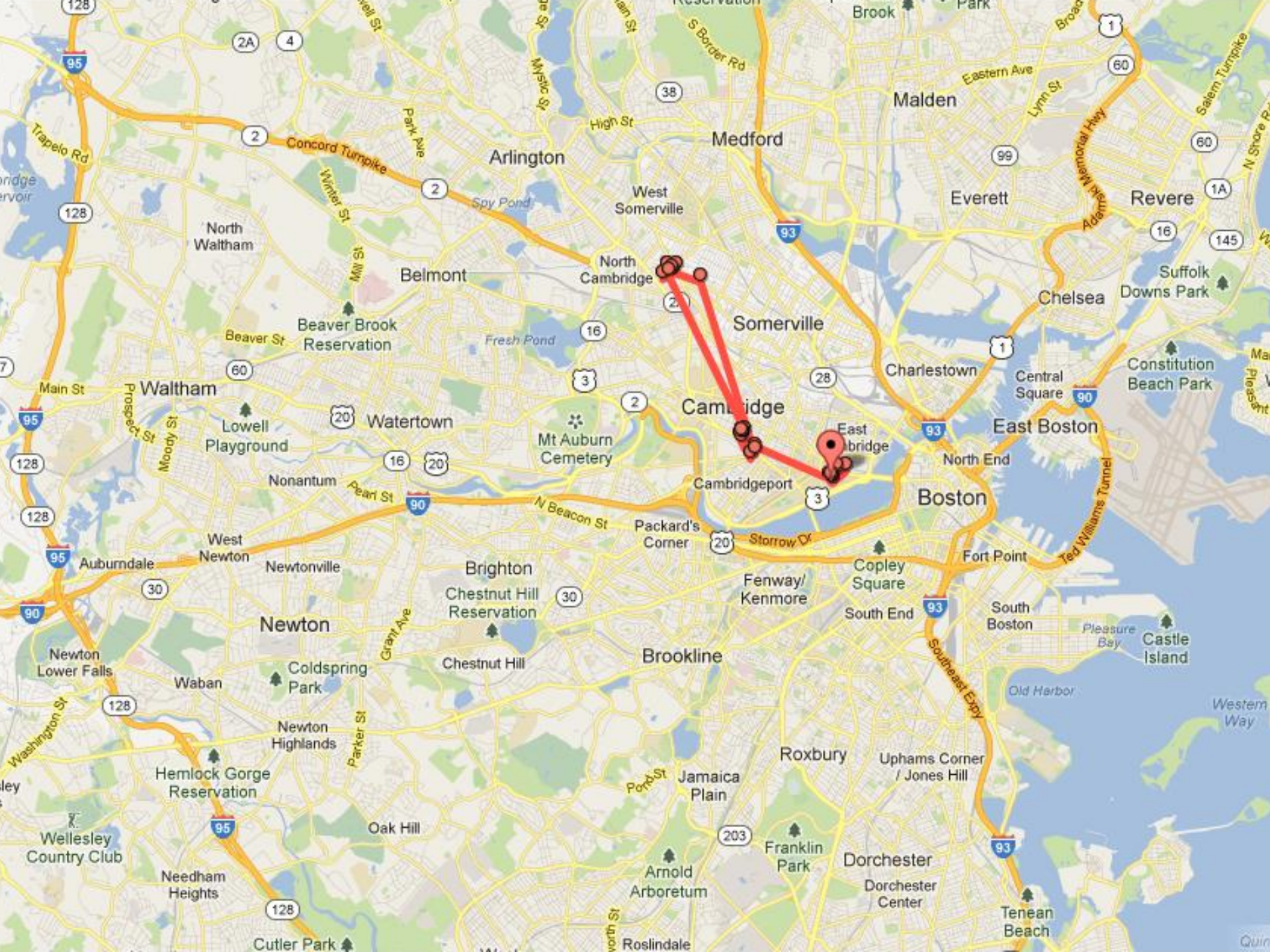

EDMOND LOCARD
TRAITÉ
DE
CRIMINALISTIQUE
TOME PREMIER
LES EMPREINTES
ET LES TRACES
DANS L'ENQUÊTE CRIMINELLE
(PREMIÈRE PARTIE)
LYON
Joannès DESVIGNE et ses Fils
Imprimeurs-Éditeurs
36 à 40, Passage de l'Hôtel-Dieu
1931
Traité de Criminalistique
Edmond LOCARD

12 points

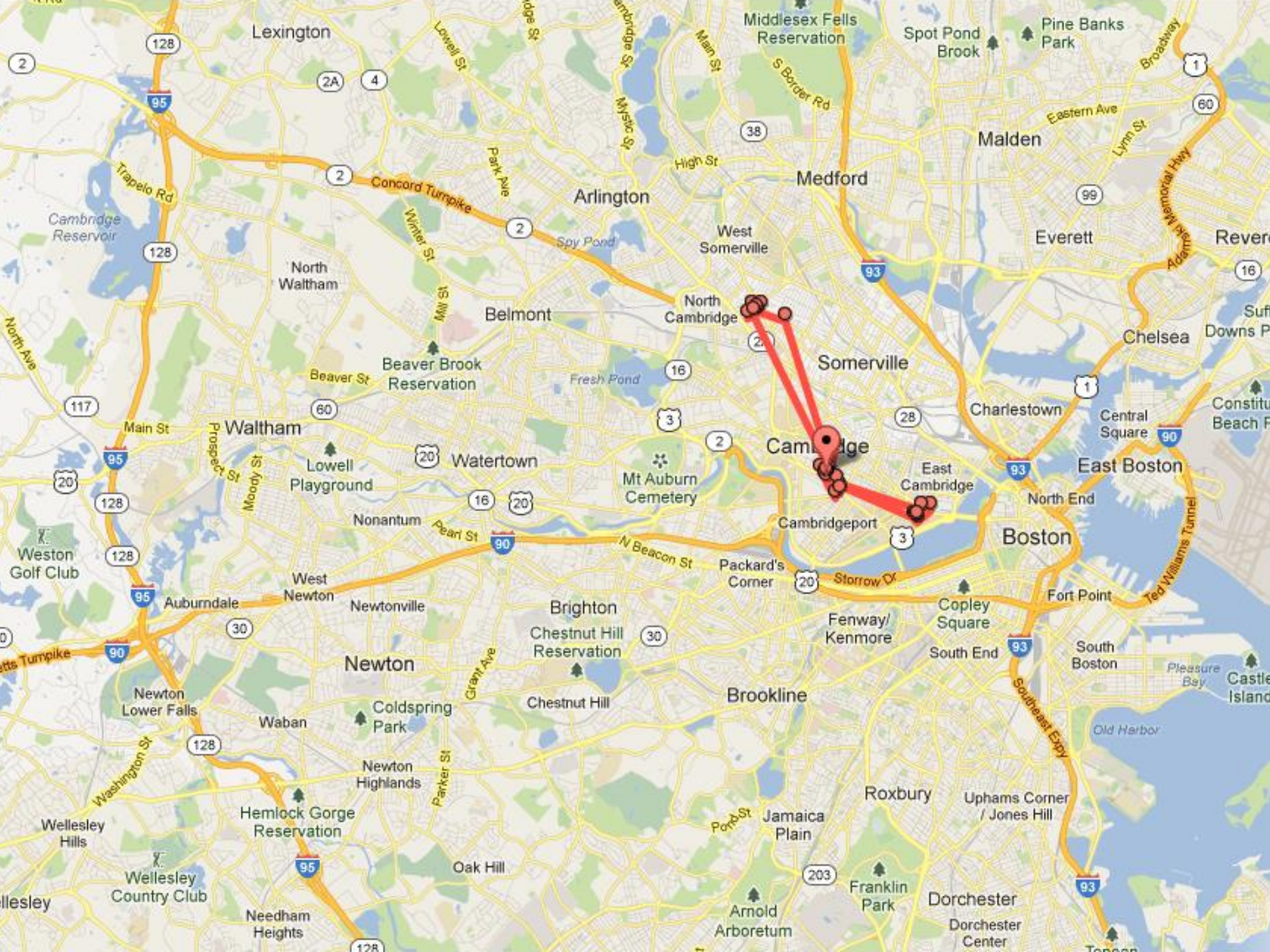


Somerville
Spy Pond
Marsh St
Concord Ave
Pleasant St
Brighton St
Concord Turnpike
Lake St
Broadway
Harvard St
Mystic River Reservation
Fellsway
Belmont
North Cambridge
Davis Square
Cedar St
Medford St
Broadway
Rindge Ave
Mill St
Waverley St
Sherman St
Walden St
Elm St
Summer St
Foss Park
Somerville
Alford St
Ryan Playground
Washington St
Fresh Pond
Beacon St
Cross St
Pequossette Playground
White St
Belmont St
Concord Ave
Huron Ave
Oxford St
Park St
Washington St
Charlestown
Orchard St
Common St
School St
Brattle St
McGrath Hwy
Thompson Square / Bunker Hill
Victory Field
Oakley Country Club
Arlington St
Mt Auburn St
Cambridge
Prospect St
Cambridge St
Watertown
Main St
Mt Auburn St
Mt Auburn Cemetery
John F Kennedy Park
Cooper Park
Broadway
Hampshire St
East Cambridge
Pleasant St
Walnut St
Greenough Blvd
Hoyt Field
Binney St
California St
Arsenal St
Western Ave
N Harvard St
Massachusetts Ave
Main St
West End
Watertown St
Galen St
N Beacon St
Charles River Rd
Cambridgeport
Charles River Basin
Pearl St
Cambridge St
Vassar St
Boston
Adams St
Massachusetts Turnpike (Toll road)
Boston Common
Faneuil St
Market St
Brighton Ave
Storrow Dr
Burr Playground
Brighton
Copley Square
Chinatown
Cabot Park
Centre St
Kenrick St
Brighton
Babcock St
Pleasant St
Corey Hill Outlook Park
Fenway/ Kenmore
Back Bay Fens
Tremont St
Foster St
Washington St
South End
Parsons Field
Riverway
Parker St
Commonwealth Avenue
Chestnut Hill Reservation
Beacon St
Tappan St
Clinton Rd
Melnea Cass Blvd
Albany St
Beacon St
Brookline
Chestnut Hill

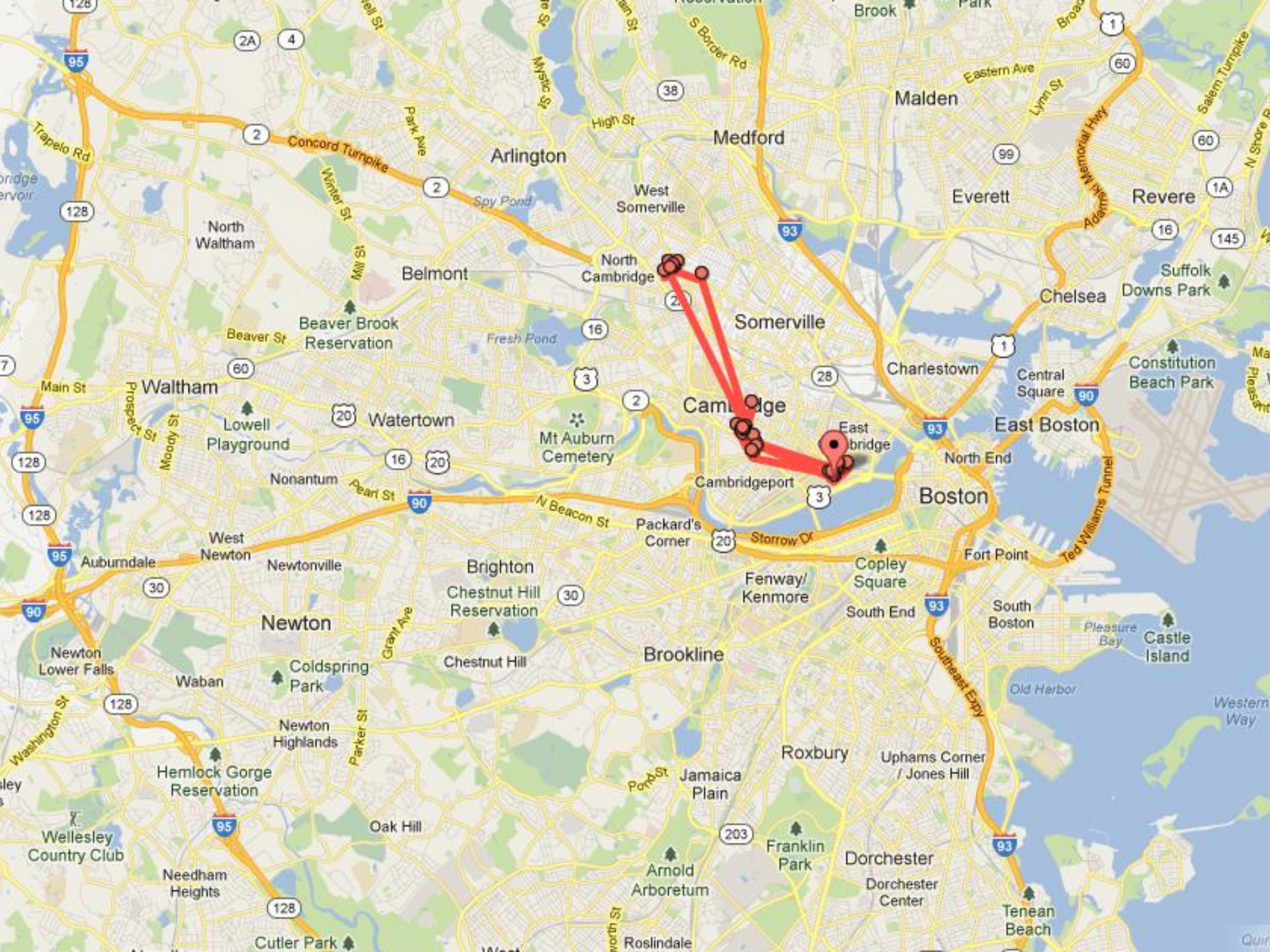

Arlington Reservoir
Playstead Park
Victory Park
Fellsmere Park
Glenwood
Medford
Devir Park
Robbins Farm Park
Arlington
Belmont Country Club
Menotomy Rocks Park
Spy Pond
West Somerville
Mystic River Reservation
Davis Square
North Cambridge
Belmont
Fresh Pond
Foss Park
Somerville
Pequossette Playground
Victory Field
Oakley Country Club
Watertown
Mt Auburn Cemetery
John F Kennedy Park
Cambridge
Cooper Park
Hoyt Field
East Cambridge
Cambridgeport
Nonantum
Charles River Basin
Concord Turnpike
Massachusetts Turnpike (Toll road)
Massachusetts Ave
Mystic Valley Pkwy
Fellsway W
Salem St
High St
Summer St
Broadway
Harvard St
Riverside Ave
Concord Ave
Pleasant St
Lake St
Rindge Ave
Huron Ave
Mt Auburn St
Washington St
Main St
Cambridge St
Waverley St
Belmont St
Greenough Blvd
Western Ave
Arsenal St
Walnut St
Charles River Rd
N Beacon St
Pearl St
Crafts St
California St
Waltham St
Seyon St
Trapelo Rd
Winter St
Marsh St
Appleton St
Bow St
Lowell St
Forest St
Washington St
Mystic St
Winthrop St
Fulton St
Governors Ave
Spring St
Fellsway
Middlesex Ave
Rivers Edge Dr
Boston Ave
College Ave
Medford St
Cedar St
Elm St
Beacon St
Park St
Prospect St
Hampshire St
McGrath Hwy
Cross St
Binney St
Sherman St
Walden St
Brattle St
Oxford St
Massachusetts Ave
Vassar St
Cambridge St
Mill St
White St
School St
Arlington St
Common St
Orchard St
Warren St
Galen St
Watertown St
Adams St
Pleasant St
Brighton St
Concord Ave
Waverley Oaks Rd
Pleasant St
Appleton St
Park Ave
Pleasant St
Massachusetts Ave
Summer St
Forest St
Washington St
Lowell St
Pleasant St
Winter St
Marsh St
Concord Ave
Trapelo Rd
Mill St
Waverley St
Belmont St
Orchard St
Warren St
Main St
Seyon St
Waltham St
Pleasant St
Main St
Mt Auburn St
Walnut St
California St
Nonantum
Watertown St
Galen St
Charles River Rd
N Beacon St
Arsenal St
Pearl St
Crafts St
Adams St
Greenough Blvd
Western Ave
Harvard St
N Harvard St
Cambridge St
Vassar St
Massachusetts Ave
Main St
Binney St
Hampshire St
Broadway
Prospect St
McGrath Hwy
Cross St
Fellsway
Middlesex Ave
Rivers Edge Dr
Fellsway W
Salem St
Riverside Ave
Harvard St
Broadway
Cambridge St
Mt Auburn St
Brattle St
Washington St
Huron Ave
Concord Ave
Rindge Ave
Lake St
Concord Turnpike
Massachusetts Ave
Mystic Valley Pkwy
High St
Summer St
Appleton St
Bow St
Massachusetts Ave
Summer St
Washington St
Mystic St
Winthrop St
Fulton St
Governors Ave
Forest St
Spring St
Fellsway
Middlesex Ave
Rivers Edge Dr
Boston Ave
College Ave
Main St
Harvard St
Broadway
Riverside Ave
Concord Ave
Pleasant St
Brighton St
Lake St
Rindge Ave
Cedar St
Medford St
Broadway
Foss Park
Somerville
Cross St
Sherman St
Walden St
Elm St
Beacon St
Park St
Oxford St
Washington St
McGrath Hwy
Prospect St
Cambridge St
Hampshire St
Binney St
Main



Brook
Park
Malden
Medford
Arlington
West Somerville
Everett
Revere
North Waltham
Belmont
North Cambridge
Somerville
Chelsea
Suffolk Downs Park
Beaver Brook Reservation
Fresh Pond
Charlestown
Central Square
Constitution Beach Park
Waltham
Watertown
Cambridge
East Boston
Lowell Playground
Mt Auburn Cemetery
North End
Nonantum
Cambridgeport
Boston
Packard's Corner
West Newton
Auburndale
Newtonville
Brighton
Copley Square
Fort Point
Chestnut Hill Reservation
Fenway/ Kenmore
South End
South Boston
Newton
Pleasure Bay
Castle Island
Newton Lower Falls
Waban
Coldspring Park
Chestnut Hill
Brookline
Old Harbor
Western Way
Newton Highlands
Roxbury
Uphams Corner / Jones Hill
Hemlock Gorge Reservation
Jamaica Plain
Oak Hill
Wellesley Country Club
Franklin Park
Dorchester
Needham Heights
Arnold Arboretum
Dorchester Center
Tenean Beach
Cutler Park
Roslindale
Concord Turnpike
Trapelo Rd
Spy Pond
Beaver St
Main St
Prospect St
Moody St
Pearl St
N Beacon St
Storrow Dr
Ted Williams Tunnel
Southeast Expy
Eastern Ave
Lynn St
Salem Turnpike
Adamski Memorial Hwy
S Border Rd
High St
Mystic St
Park Ave
Winter St
Mill St
Grant Ave
Parker St
Washington St
Pond St



Lexington
Middlesex Fells Reservation
Spot Pond Brook
Pine Banks Park
Malden
Medford
Arlington
Concord Turnpike
Trapelo Rd
Cambridge Reservoir
Spy Pond
West Somerville
Everett
North Waltham
Belmont
North Cambridge
Somerville
Chelsea
Beaver Brook Reservation
Fresh Pond
Charlestown
Central Square
Waltham
Lowell Playground
Watertown
Mt Auburn Cemetery
East Cambridge
East Boston
North End
Nonantum
Cambridgeport
Boston
Weston Golf Club
Packard's Corner
Storrow Dr
West Newton
Auburndale
Newtonville
Brighton
Copley Square
Fort Point
Chestnut Hill Reservation
Fenway/ Kenmore
South End
South Boston
Newton
Newton Lower Falls
Waban
Coldspring Park
Chestnut Hill
Brookline
Old Harbor
Newton Highlands
Roxbury
Uphams Corner / Jones Hill
Hemlock Gorge Reservation
Jamaica Plain
Wellesley Hills
Oak Hill
Wellesley Country Club
Franklin Park
Dorchester
Needham Heights
Arnold Arboretum
Dorchester Center



Brook
Park
Malden
Medford
Arlington
Concord Turnpike
Trapelo Rd
Spy Pond
West Somerville
Everett
Revere
North Waltham
Belmont
North Cambridge
Somerville
Chelsea
Suffolk Downs Park
Beaver Brook Reservation
Beaver St
Fresh Pond
Charlestown
Central Square
Constitution Beach Park
Waltham
Main St
Lowell Playground
Watertown
Mt Auburn Cemetery
East Boston
North End
Nonantum
Pearl St
Cambridgeport
Boston
N Beacon St
Packard's Corner
Storrow Dr
West Newton
Auburndale
Newtonville
Brighton
Copley Square
Fort Point
Fenway/ Kenmore
Chestnut Hill Reservation
South End
South Boston
Newton
Pleasure Bay
Castle Island
Newton Lower Falls
Waban
Coldspring Park
Chestnut Hill
Brookline
Southeast Expy
Old Harbor
Western Way
Newton Highlands
Roxbury
Uphams Corner / Jones Hill
Hemlock Gorge Reservation
Jamaica Plain
Oak Hill
Wellesley Country Club
Franklin Park
Dorchester
Needham Heights
Arnold Arboretum
Dorchester Center
Tenean Beach
Cutler Park
Roslindale

Is the way you move
as **unique** as
your fingerprint

We can use points to identify a fingerprint

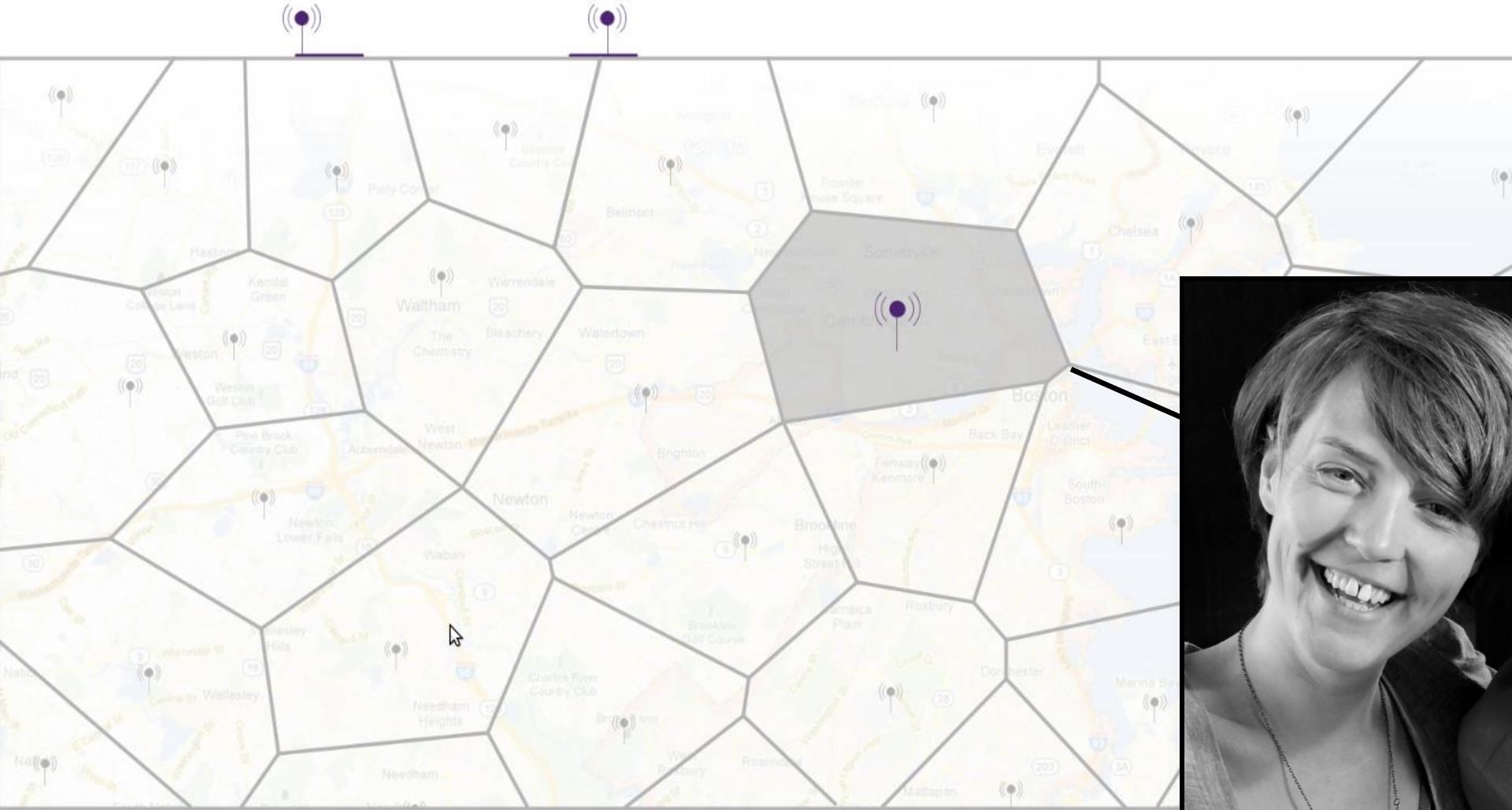
Scott

1 point for mobility data

From 10 to 11am

2 points

Around 11:30am



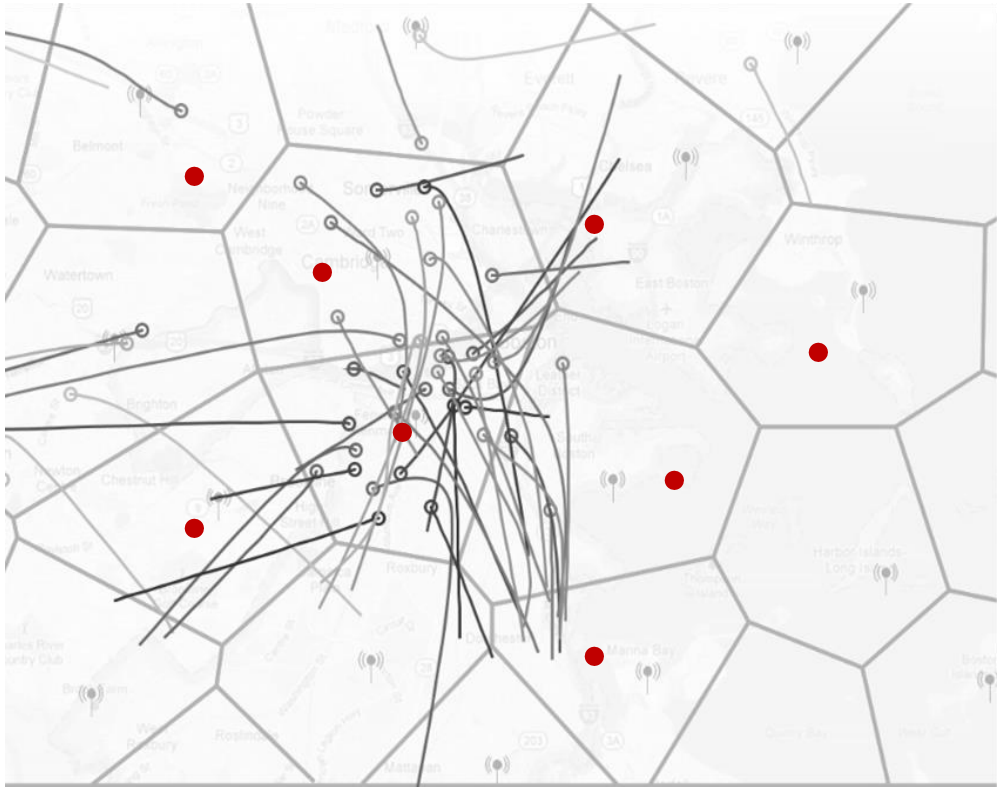
3 points

For lunch

Video 1

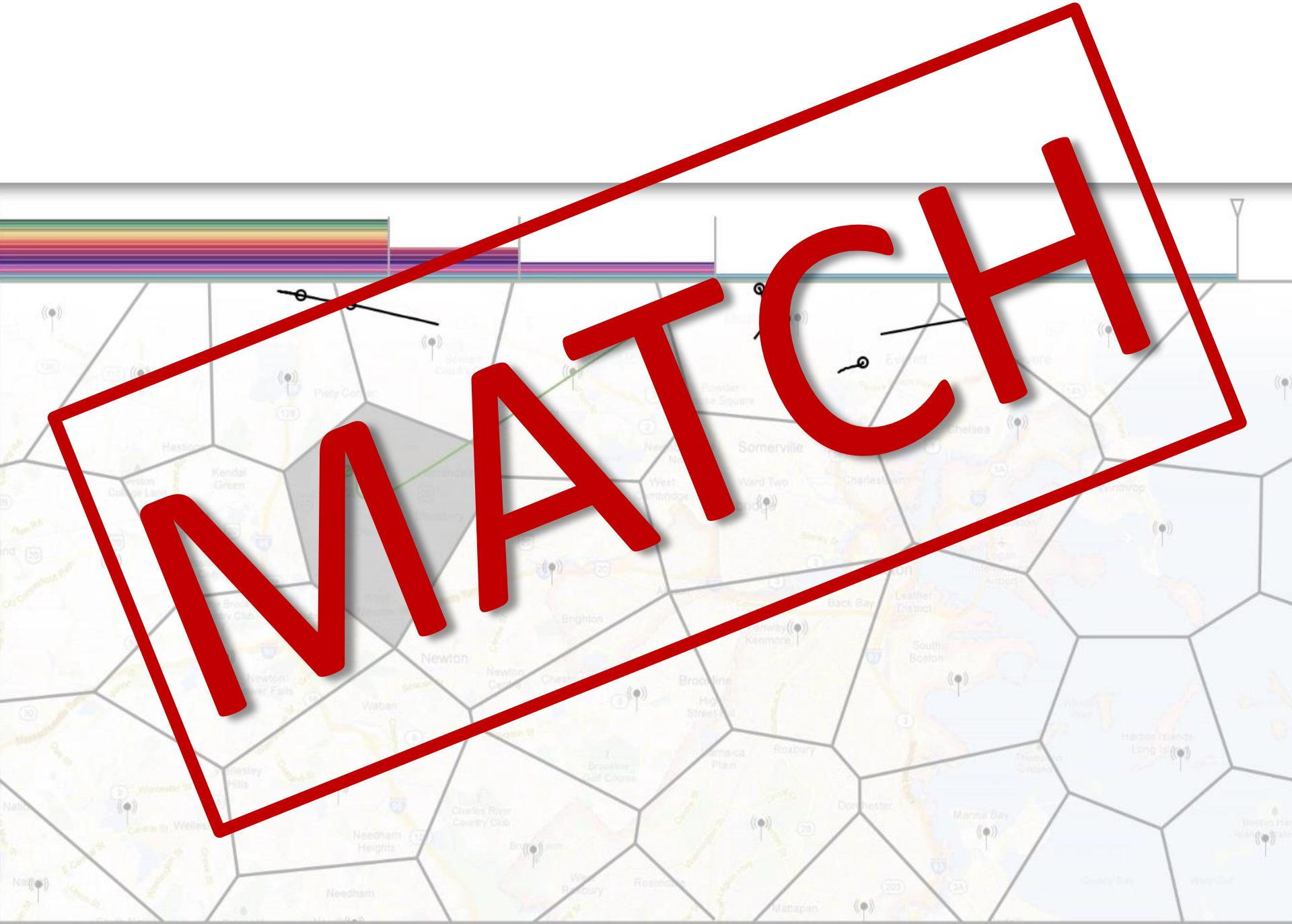
Video1

How many points do I need to uniquely identify a mobility traces?



Video 2

Video2

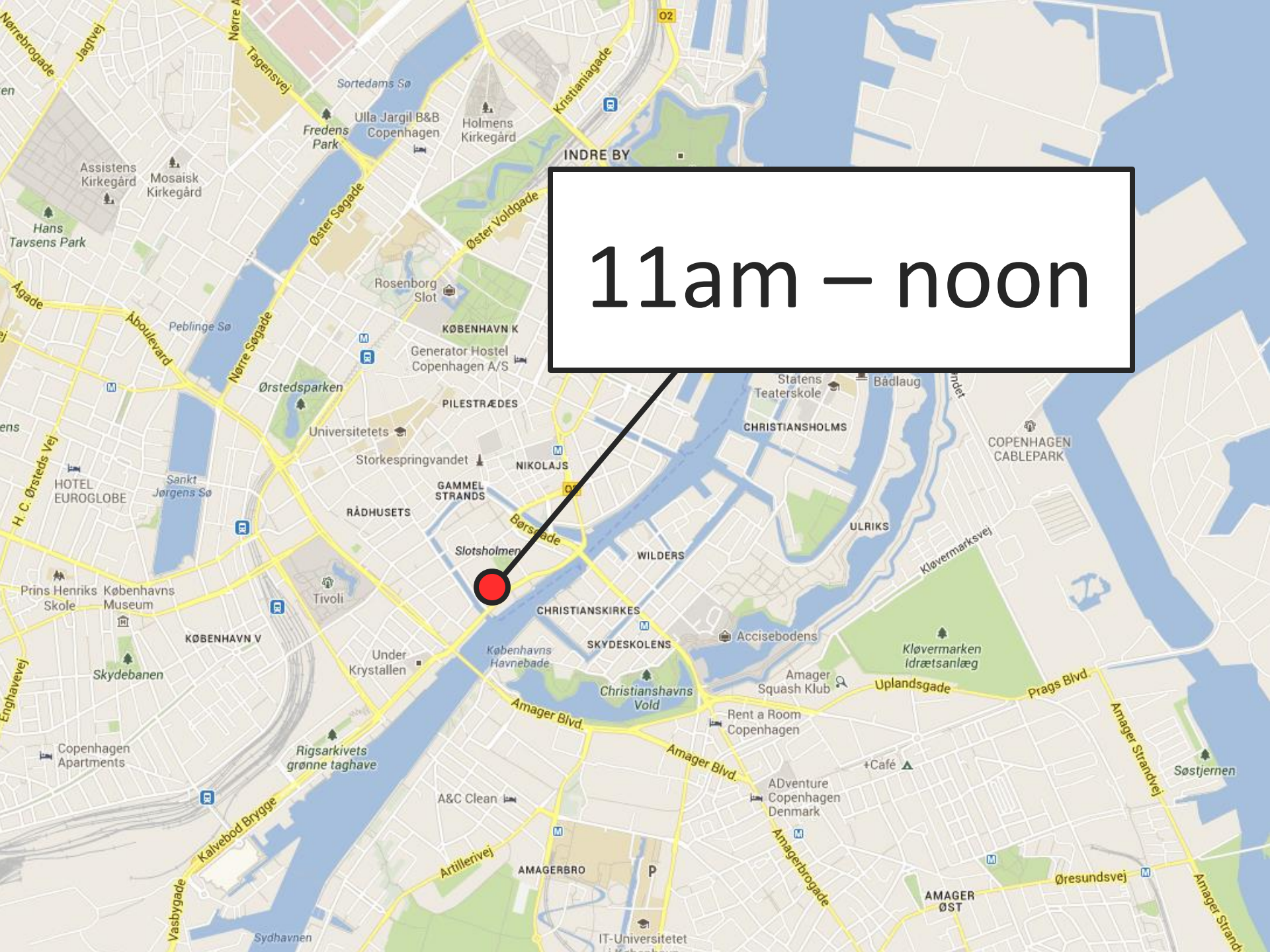

MATCH

Entire country of 1.5 millions people

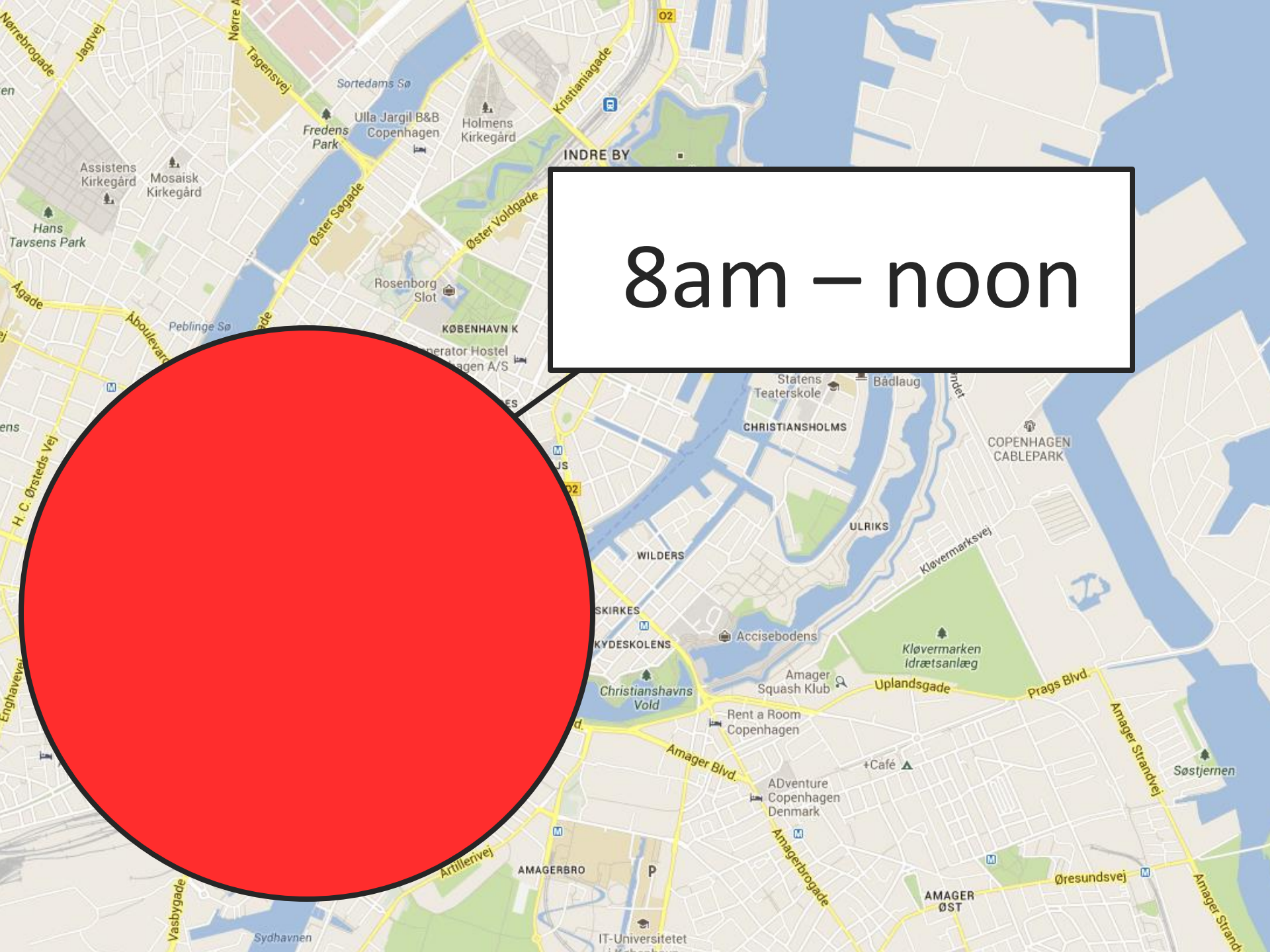
Our behavior is unique enough

4 points

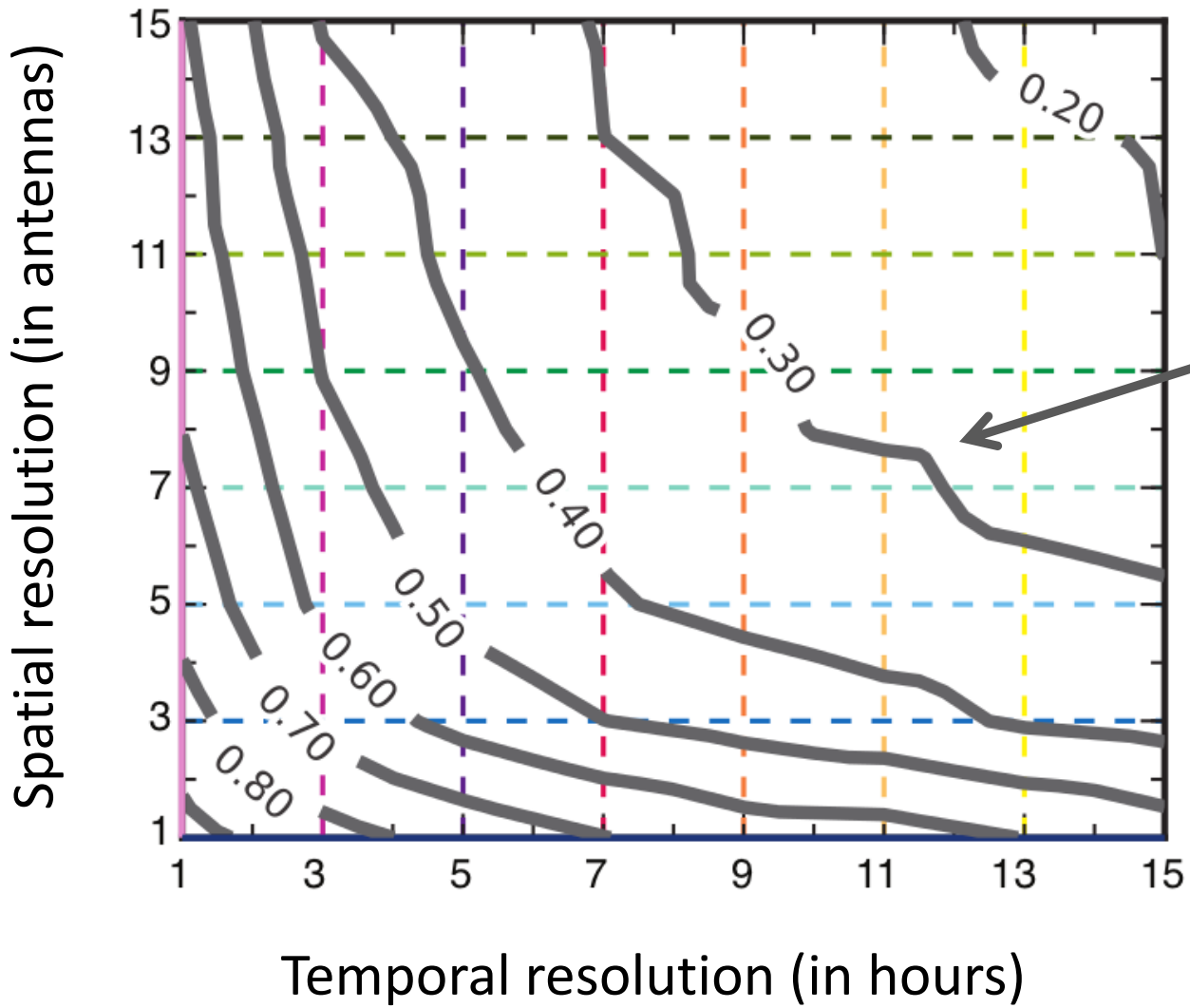
Identify 95% of people


11am – noon
Nørrebrogade
Jagtvej
Nørre Allé
Tagensvej
Sortedams Sø
Fredens Park
Ulla Jargil B&B Copenhagen
Holmens Kirkegård
Kristianiagade
INDRE BY
Assistens Kirkegård
Mosaisk Kirkegård
Hans Tavsens Park
Øster Søgade
Øster Voldgade
Rosenborg Slot
Ågade
Åboulevard
Peblinge Sø
Nørre Søgade
KØBENHAVN K
Generator Hostel Copenhagen A/S
Ørstedsparken
PILESTRÆDES
Universitetets
Statens Teaterskole
Bådlaug
CHRISTIANSHOLMS
COPENHAGEN CABLEPARK
H. C. Ørsteds Vej
HOTEL EUROGLOBE
Sankt Jørgens Sø
Storkespringvandet
NIKOLAJS
GAMMEL STRANDS
RÅDHUSETS
Børsgade
Slotsholmen
WILDERS
ULRIKS
Kløvermarksvej
Prins Henriks Skole
Københavns Museum
Tivoli
CHRISTIANSKIRKES
KØBENHAVN V
SKYDESKOLENS
Accisebodens
Kløvermarken Idrætsanlæg
Under Krystallen
Københavns Havnebade
Skydebanen
Christianshavns Vold
Amager Squash Klub
Uplandsgade
Prags Blvd.
Amager Blvd.
Rent a Room Copenhagen
Amager Strandvej
Copenhagen Apartments
Rigsarkivets grønne taghave
+Café
Søstjernen
ADventure Copenhagen Denmark
A&C Clean
Kalvebod Brygge
Artillerivej
AMAGERBRO
Amagerbrogade
Øresundsvej
AMAGER ØST
Vasbygade
Sydhavnen
IT-Universitetet

8am – noon

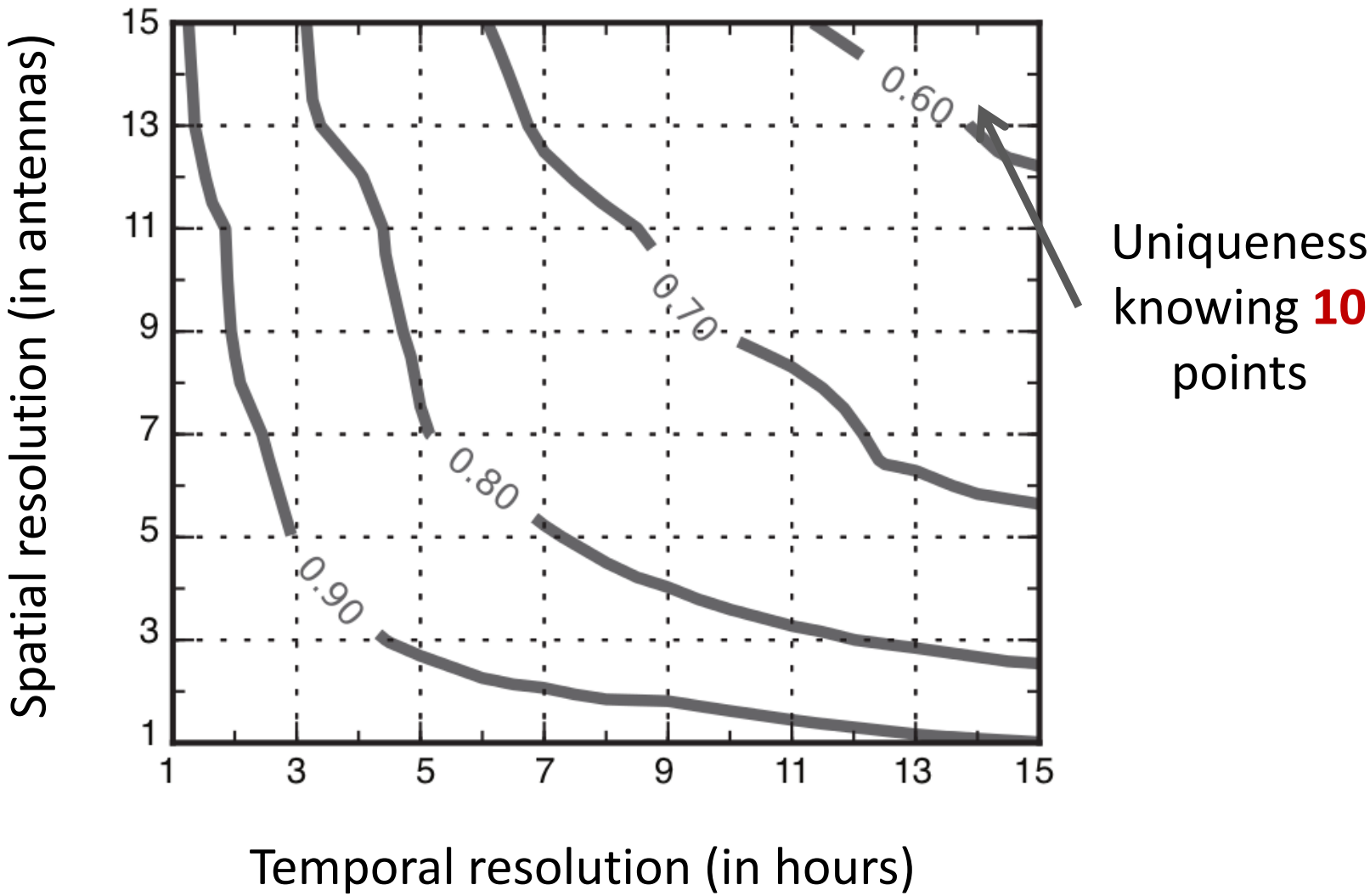

Nørrebrogade
Jagtvej
Nørre A
Tagensvej
Sortedams Sø
Fredens Park
Ulla Jargil B&B Copenhagen
Holmens Kirkegård
Kristianiagade
INDRE BY
Assistens Kirkegård
Mosaisk Kirkegård
Hans Tavsens Park
Øster Søgade
Øster Voldgade
Rosenborg Slot
Ågade
Åboulevard
Peblinge Sø
KØBENHAVN K
Statens Teaterskole
Bådlaug
CHRISTIANSHOLMS
COPENHAGEN CABLEPARK
H. C. Ørsteds Vej
ULRIKS
WILDERS
Kløvermarksvej
SKIRKES
KYDESKOLENS
Accisebodens
Kløvermarken Idrætsanlæg
Christianshavns Vold
Amager Squash Klub
Uplandsgade
Prags Blvd.
Rent a Room Copenhagen
Amager Blvd.
+Café
ADventure Copenhagen Denmark
Amager Strandvej
Søstjernen
Amagerbrogade
Artillerivej
AMAGERBRO
Øresundsvej
AMAGER ØST
Vasbygade
Sydhavnen
IT-Universitetet

What about resolution?

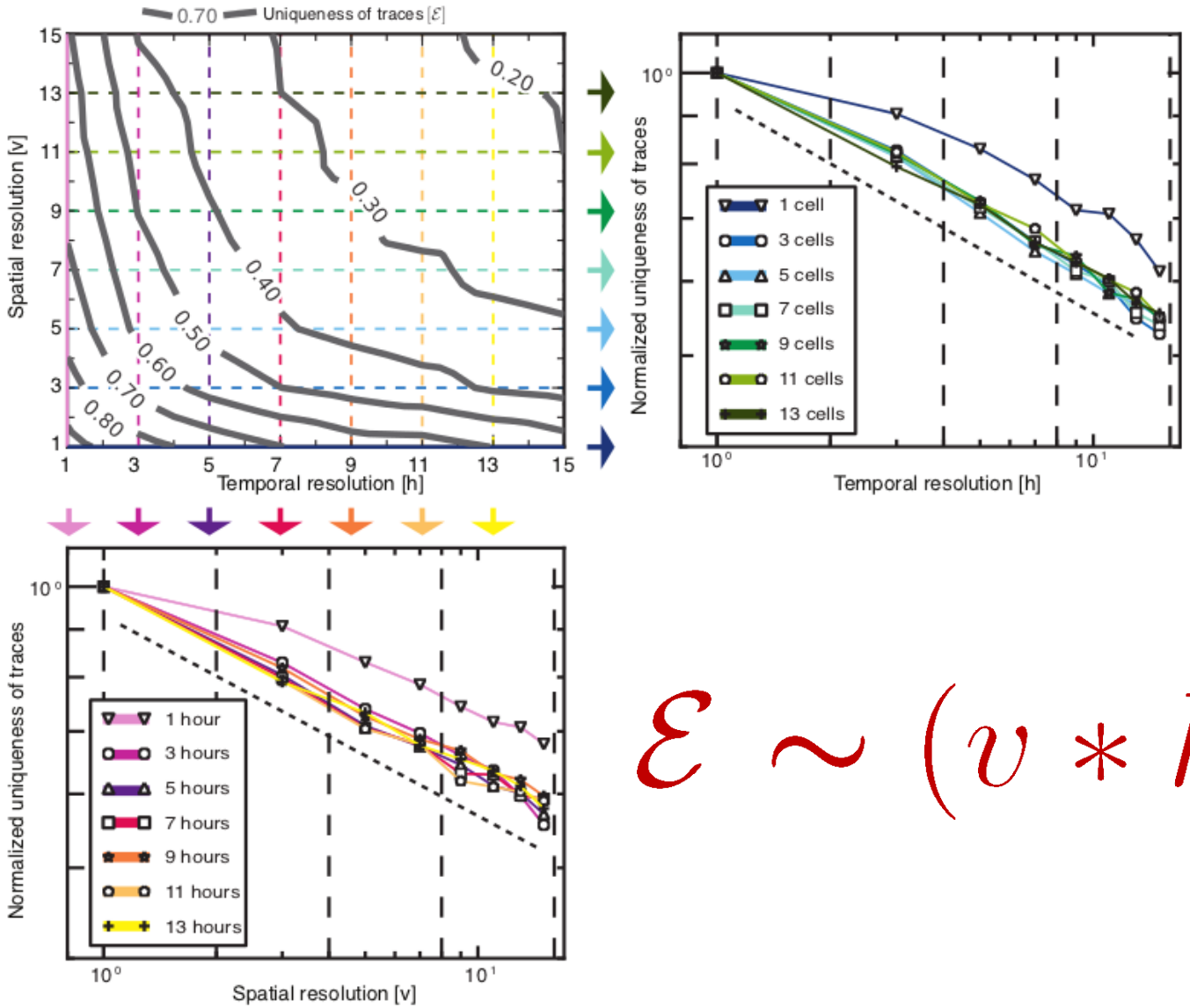


Uniqueness knowing **4** points

What about resolution but when we know 10 points



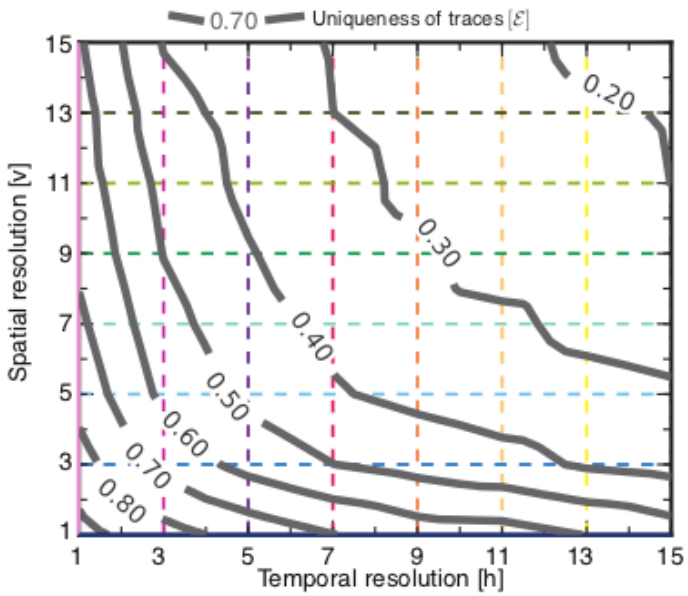
It is getting harder and harder to "gain" privacy

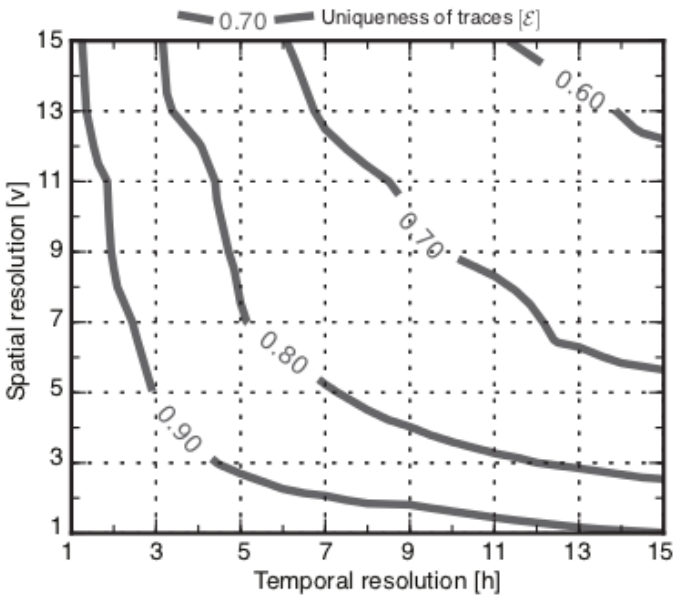


$$\mathcal{E} \sim (v * h)^{\beta}$$

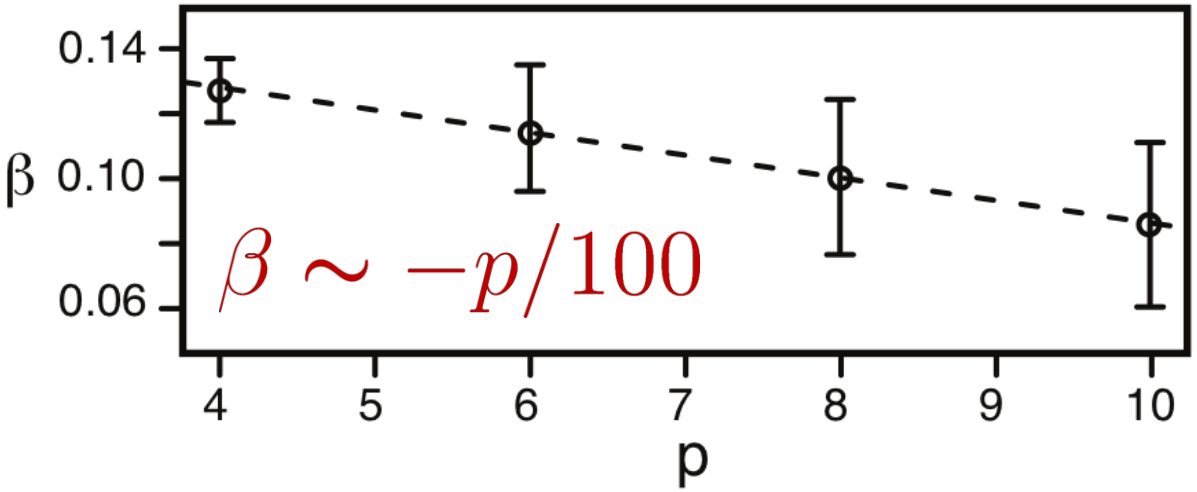
Number of points

4 points





10 points



Unicity

Number of points

$$\mathcal{E} \sim (v * h)^{-p/100}$$

Spatial resolution

Temporal resolution

It’s hard to hide in the crowd

Unicity: quantifying the privacy-utility trade-off

Simple anonymization does not work even when the data is coarse

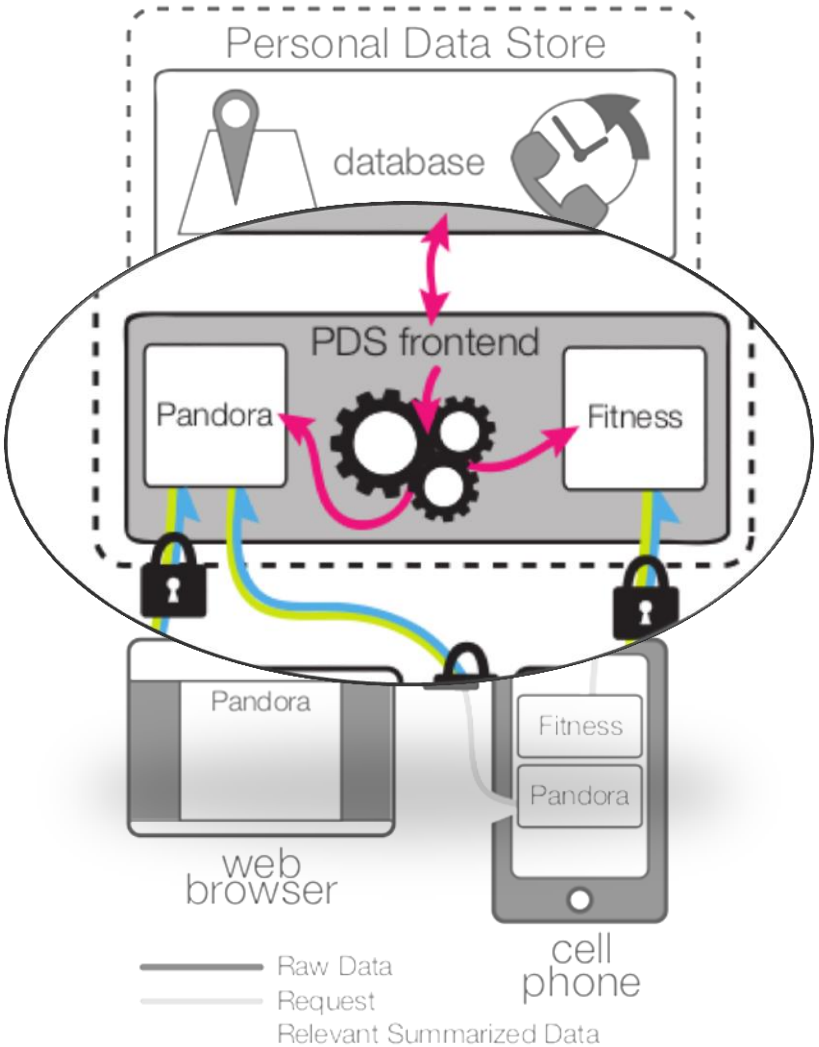
Smarter anonymization

- Application specific
- Assume that we anonymize and release the data once and for all

Online systems

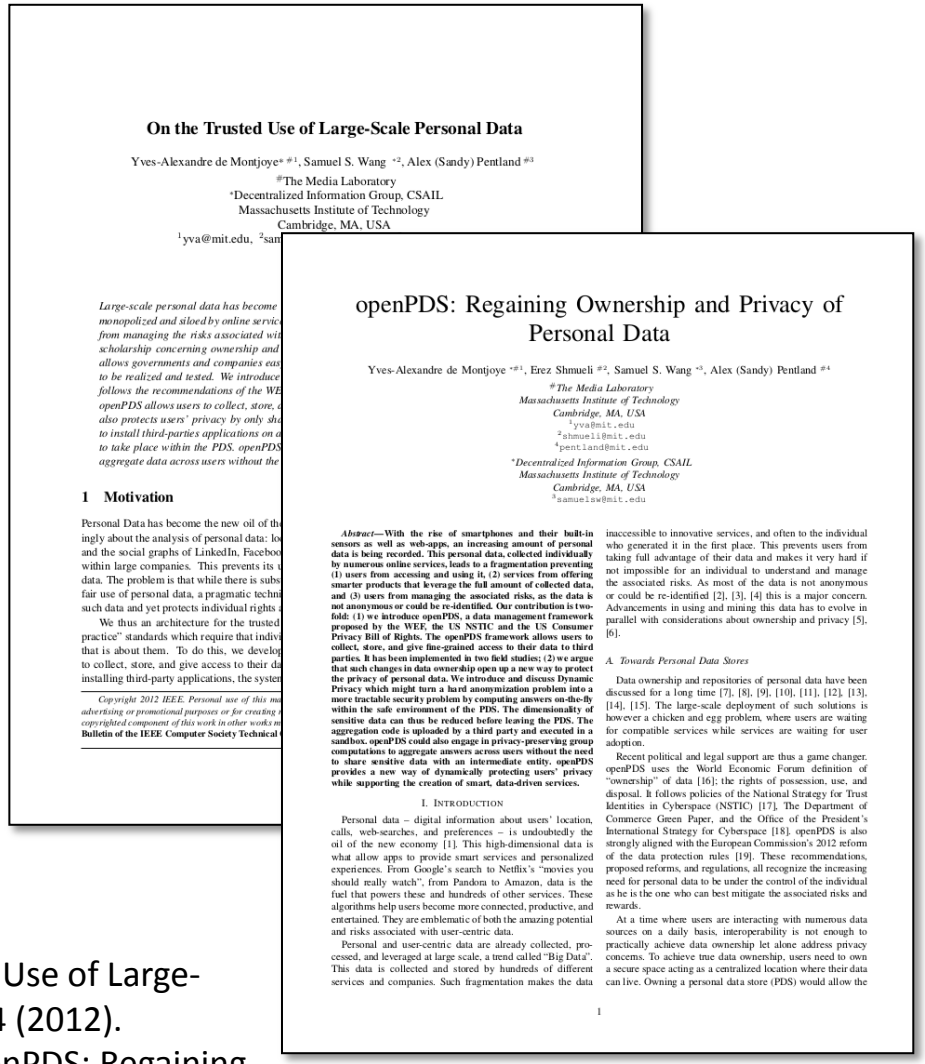
- Access to individual mobility data is a "side effect"
- Answers are what we're interested in

Online systems: from privacy to security using openPDS



- Only shares answers, not raw data
- Individual quantification of the risks

It is already happening!



On the Trusted Use of Large-Scale Personal Data

Yves-Alexandre de Montjoye* #1, Samuel S. Wang *2, Alex (Sandy) Pentland #3
#The Media Laboratory
*Decentralized Information Group, CSAIL
Massachusetts Institute of Technology
Cambridge, MA, USA
1yva@mit.edu, 2sam

Large-scale personal data has become
monopolized and siloed by online servic
from managing the risks associated wit
scholarship concerning ownership and
allows governments and companies eas
to be realized and tested. We introduce
follows the recommendations of the WE
openPDS allows users to collect, store,
also protects users' privacy by only sha
to install third-parties applications on a
to take place within the PDS. openPDS
aggregate data across users without the

1 Motivation

Personal Data has become the new oil of th
ingly about the analysis of personal data: lo
and the social graphs of LinkedIn, Faceboo
within large companies. This prevents its u
data. The problem is that while there is subs
fair use of personal data, a pragmatic techni
such data and yet protects individual rights

We thus an architecture for the trusted
practice" standards which require that indiv
that is about them. To do this, we develop
to collect, store, and give access to their da
installing third-party applications, the system

Copyright 2012 IEEE. Personal use of this ma
advertising or promotional purposes or for creating
copyrighted component of this work in other works m
Bulletin of the IEEE Computer Society Technical

openPDS: Regaining Ownership and Privacy of Personal Data

Yves-Alexandre de Montjoye *#1, Erez Shmueli #2, Samuel S. Wang *3, Alex (Sandy) Pentland #4
#The Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA, USA
1yva@mit.edu
2shmueli@mit.edu
4pentland@mit.edu
**Decentralized Information Group, CSAIL*
Massachusetts Institute of Technology
Cambridge, MA, USA
3samuelsw@mit.edu

***Abstract*—With the rise of smartphones and their built-in sensors as well as web-apps, an increasing amount of personal data is being recorded. This personal data, collected individually by numerous online services, leads to a fragmentation preventing (1) users from accessing and using it, (2) services from offering smarter products that leverage the full amount of collected data, and (3) users from managing the associated risks, as the data is not anonymous or could be re-identified. Our contribution is two-fold: (1) we introduce openPDS, a data management framework proposed by the WEF, the US NSTIC and the US Consumer Privacy Bill of Rights. The openPDS framework allows users to collect, store, and give fine-grained access to their data to third parties. It has been implemented in two field studies; (2) we argue that such changes in data ownership open up a new way to protect the privacy of personal data. We introduce and discuss Dynamic Privacy which might turn a hard anonymization problem into a more tractable security problem by computing answers on-the-fly within the safe environment of the PDS. The dimensionality of sensitive data can thus be reduced before leasing the PDS. The aggregation code is uploaded by a third party and executed in a sandbox. openPDS could also engage in privacy-preserving group computations to aggregate answers across users without the need to share sensitive data with an intermediate entity. openPDS provides a new way of dynamically protecting users' privacy while supporting the creation of smart, data-driven services.**

I. Introduction

Personal data – digital information about users' location, calls, web-searches, and preferences – is undoubtedly the oil of the new economy [1]. This high-dimensional data is what allow apps to provide smart services and personalized experiences. From Google's search to Netflix's "movies you should really watch", from Pandora to Amazon, data is the fuel that powers these and hundreds of other services. These algorithms help users become more connected, productive, and entertained. They are emblematic of both the amazing potential and risks associated with user-centric data.

Personal and user-centric data are already collected, processed, and leveraged at large scale, a trend called "Big Data". This data is collected and stored by hundreds of different services and companies. Such fragmentation makes the data inaccessible to innovative services, and often to the individual who generated it in the first place. This prevents users from taking full advantage of their data and makes it very hard if not impossible for an individual to understand and manage the associated risks. As most of the data is not anonymous or could be re-identified [2], [3], [4] this is a major concern. Advancements in using and mining this data has to evolve in parallel with considerations about ownership and privacy [5], [6].

A. Towards Personal Data Stores

Data ownership and repositories of personal data have been discussed for a long time [7], [8], [9], [10], [11], [12], [13], [14], [15]. The large-scale deployment of such solutions is however a chicken and egg problem, where users are waiting for compatible services while services are waiting for user adoption.

Recent political and legal support are thus a game changer. openPDS uses the World Economic Forum definition of "ownership" of data [16]; the rights of possession, use, and disposal. It follows policies of the National Strategy for Trust Identities in Cyberspace (NSTIC) [17], The Department of Commerce Green Paper, and the Office of the President's International Strategy for Cyberspace [18]. openPDS is also strongly aligned with the European Commission's 2012 reform of the data protection rules [19]. These recommendations, proposed reforms, and regulations, all recognize the increasing need for personal data to be under the control of the individual as he is the one who can best mitigate the associated risks and rewards.

At a time where users are interacting with numerous data sources on a daily basis, interoperability is not enough to practically achieve data ownership let alone address privacy concerns. To achieve true data ownership, users need to own a secure space acting as a centralized location where their data can live. Owning a personal data store (PDS) would allow the

1

de Montjoye Y.-A., Wang S., Pentland A., On the Trusted Use of Large-Scale Personal Data. IEEE Data Engineering Bulletin, 35-4 (2012).
de Montjoye Y.-A., Shmueli E., Wang S., Pentland A., openPDS: Regaining Ownership and Privacy of Personal Data, Submitted

Yves-Alexandre de Montjoye
MIT Media Lab
yva@mit.edu
http://deMontjoye.com

In collaboration with Alex “Sandy” Pentland, César Hidalgo, Vincent Blondel, Michel Verleysen, Erez Shmueli, Arek Stopczynski, Sune Lehmann