Privacy and Transparency for Decision Making

Simone Fischer-Hübner

Karlstad University, Sweden

MDAI 2015

CLOUD ACCOUNTABILITY PROJECT

SmartSociety

Content

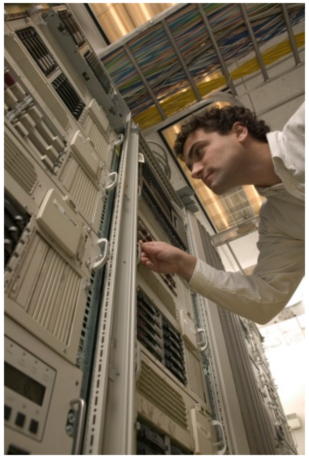
I. Profiling, Big Data & Decision Making - Privacy Challenges

II. Peer Profiling & Privacy in Smart Society

III. Transparency-enhancing Tools (TETs)

IV. Conclusions

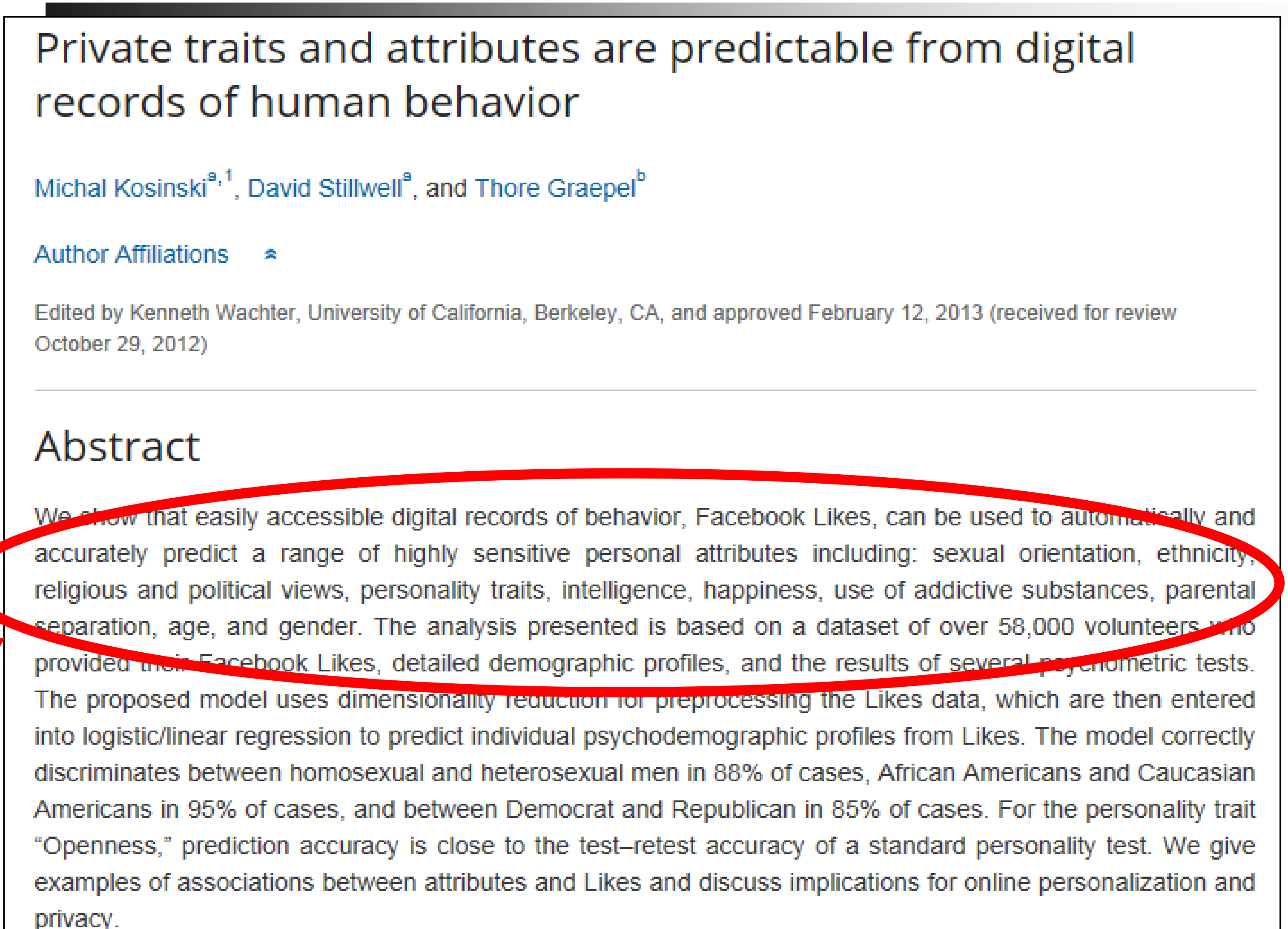
I. Profiling, Big Data & Decision Making: Privacy Risks & Challenges of Big Data



- The sheer scale of data collection, tracking, profiling & detail of the data, from many different sources
- Security of data
- Transparency
- Inaccuracy, discrimination, exclusion & economic imbalance
- Increased possibilities of government surveillance

(Art. 29 WP - Opinion 03/2013 on purpose limitation)

New types of derived data



Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

Author Affiliations

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

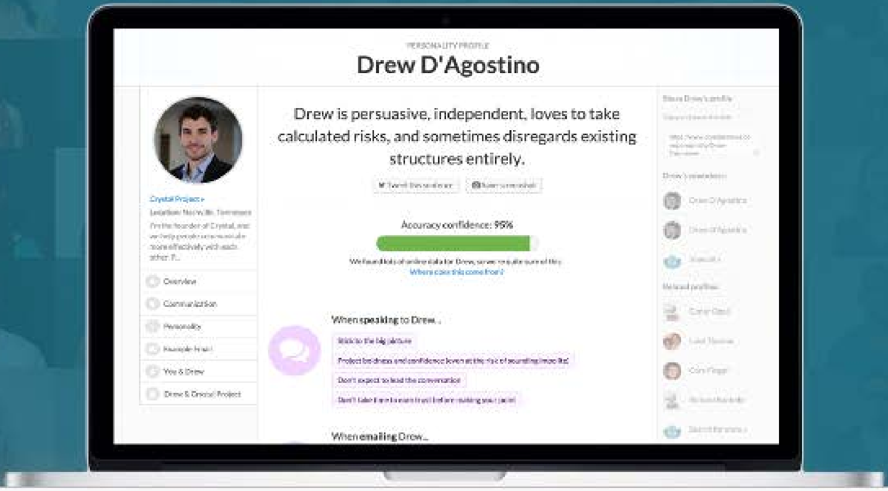
Abstract

We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented is based on a dataset of over 58,000 volunteers who provided their Facebook Likes, detailed demographic profiles, and the results of several psychometric tests. The proposed model uses dimensionality reduction for preprocessing the Likes data, which are then entered into logistic/linear regression to predict individual psychodemographic profiles from Likes. The model correctly discriminates between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. For the personality trait "Openness," prediction accuracy is close to the test–retest accuracy of a standard personality test. We give examples of associations between attributes and Likes and discuss implications for online personalization and privacy.

Crystal

SIGN UP PRICING LOG IN

Communicate with anyone ⍰ based on **personality**.



PERSONALITY PROFILE

Drew D'Agostino

Drew is persuasive, independent, loves to take calculated risks, and sometimes disregards existing structures entirely.

Accuracy confidence: 95%

When speaking to Drew...

- Stick to the big picture
- Project boldness and confidence (even at the risk of sounding imposing)
- Don't expect to lead the conversation
- Don't take time to earn trust before making your point

When emailing Drew...

Crystal shows you the best way to communicate with any coworker, prospect, or customer based on their unique personality.

Click here to try Crystal »

You'll get unlimited access to millions of personality profiles and a⍰ two-week free trial of Crystal for Gmail. No credit card required.

⍰

⍰

"If there was an award for the app that has the biggest positive impact on my work, it would go to Crystal."

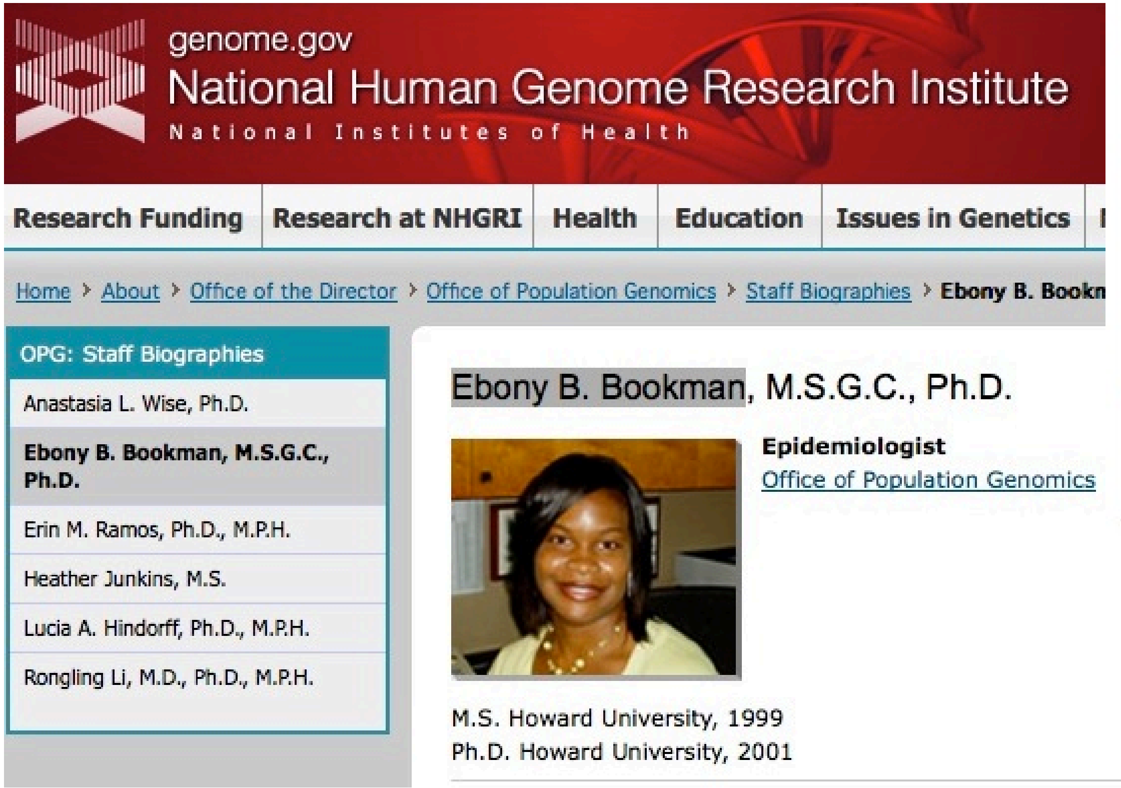
Richard Banfield, CEO of Fresh Tilled Soil

⍰

What's the difference between a bad communicator and a good one? **Empathy.**

Crystal creates **unique personality profiles** for every person with an online presence, preparing you to speak or write in someone else's natural communication style.

Personal Discrimination in Online Ad Delivery



genome.gov
National Human Genome Research Institute
National Institutes of Health

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | I

Home › About › Office of the Director › Office of Population Genomics › Staff Biographies › Ebony B. Bookn

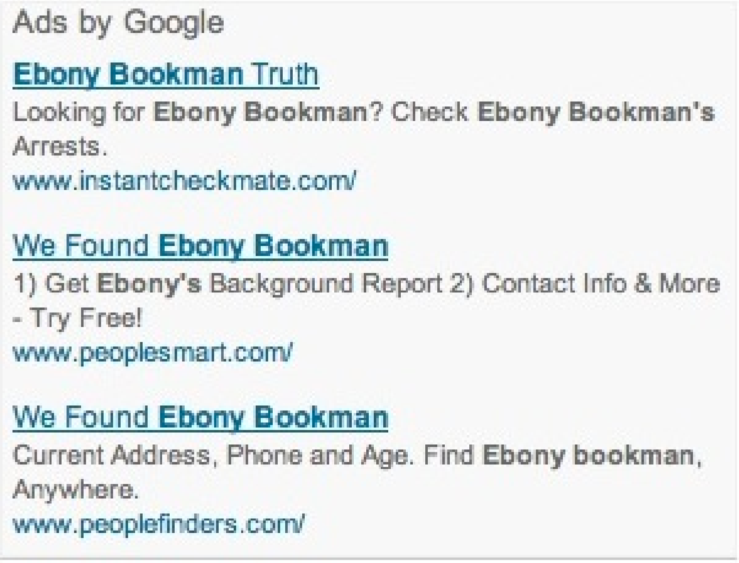
OPG: Staff Biographies

- Anastasia L. Wise, Ph.D.
- **Ebony B. Bookman, M.S.G.C., Ph.D.**
- Erin M. Ramos, Ph.D., M.P.H.
- Heather Junkins, M.S.
- Lucia A. Hindorff, Ph.D., M.P.H.
- Rongling Li, M.D., Ph.D., M.P.H.

Ebony B. Bookman, M.S.G.C., Ph.D.

Epidemiologist
Office of Population Genomics

M.S. Howard University, 1999
Ph.D. Howard University, 2001



Ads by Google

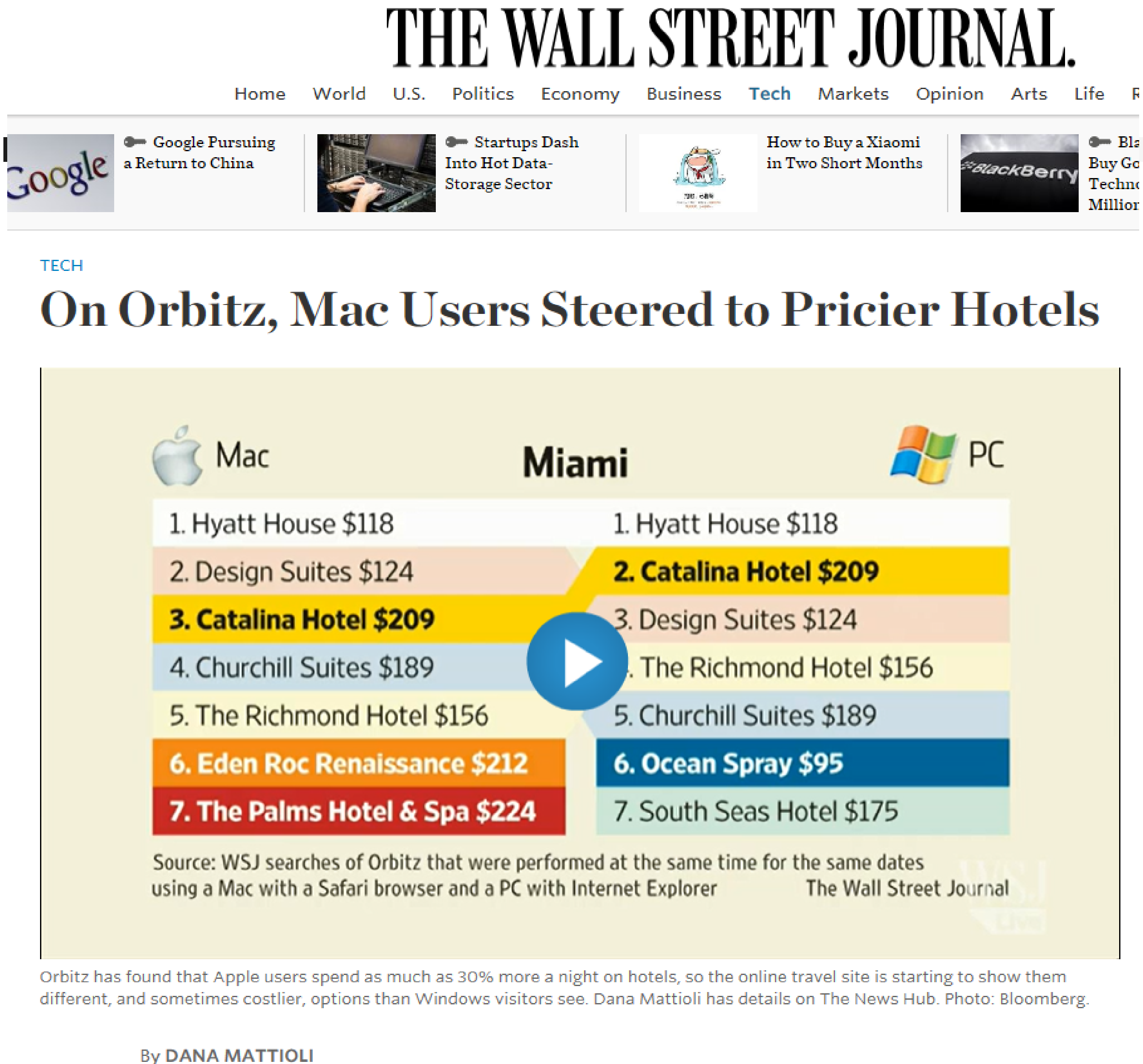
Ebony Bookman Truth
Looking for **Ebony Bookman**? Check **Ebony Bookman's** Arrests.
www.instantcheckmate.com/

We Found **Ebony Bookman**
1) Get **Ebony's** Background Report 2) Contact Info & More - Try Free!
www.peoplesmart.com/

We Found **Ebony Bookman**
Current Address, Phone and Age. Find **Ebony bookman**, Anywhere.
www.peoplefinders.com/

Source: Latanya Sweeney, dataprivacylab.org

Price Discrimination



THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life R

Google Pursuing a Return to China

Startups Dash Into Hot Data-Storage Sector

How to Buy a Xiaomi in Two Short Months

Bla
Buy Go
Techno
Millior

TECH

On Orbitz, Mac Users Steered to Pricier Hotels

Mac	PC
1. Hyatt House $118	1. Hyatt House $118
2. Design Suites $124	**2. Catalina Hotel $209**
3. Catalina Hotel $209	3. Design Suites $124
4. Churchill Suites $189	4. The Richmond Hotel $156
5. The Richmond Hotel $156	5. Churchill Suites $189
6. Eden Roc Renaissance $212	**6. Ocean Spray $95**
7. The Palms Hotel & Spa $224	7. South Seas Hotel $175

Source: WSJ searches of Orbitz that were performed at the same time for the same dates using a Mac with a Safari browser and a PC with Internet Explorer The Wall Street Journal

Orbitz has found that Apple users spend as much as 30% more a night on hotels, so the online travel site is starting to show them different, and sometimes costlier, options than Windows visitors see. Dana Mattioli has details on The News Hub. Photo: Bloomberg.

By DANA MATTIOLI

What makes the further use of personal data for analytics compatible?
(Art. 29 WP - *Opinion 03/2013 on purpose limitation*)

- If big data is analysed to detect general trends:
  - Functional separation, security & confidentiality
- If processing of big data affects individuals:
  - Opt-in consent
  - Data subject rights (transparency & intervenability) in regard to profiles and algorithms
  - Purpose limitation
  - Data minimisation
  - ..

II. Smart Society & Privacy (EU FP7 FET IP)

ICT and society today

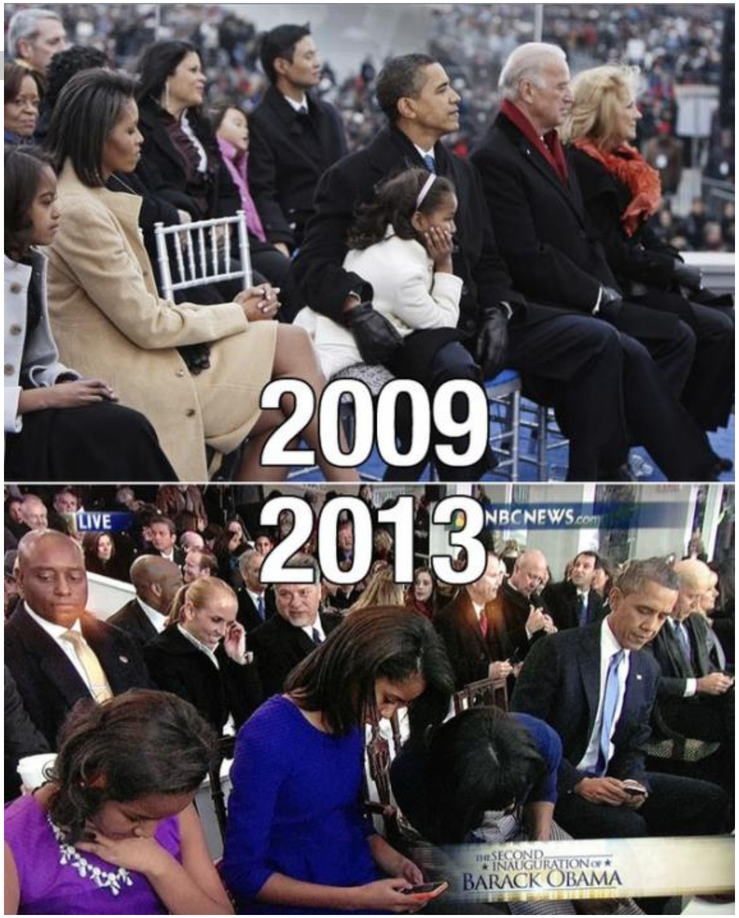
An ongoing accelerating deep integration

- Exponential increase in number of devices (sensors, PDAs, tablets, iWatch, ...)
- Machines and humans interacting more and more closely
- Physical and virtual dimensions of life are more and more intertwined

Lots of ad-hoc solutions and systems

- Social networking platforms (eg., Facebook)
- Chat systems (eg., Whatsapp)
- Mobile services anywhere anytime
- ...
- Smart city applications **(!!!)**

What about building a smarter society?



From: **When Mobile is opposite of Social**

www.smart-society-project.eu,
Source: Fausto Giunchiglia

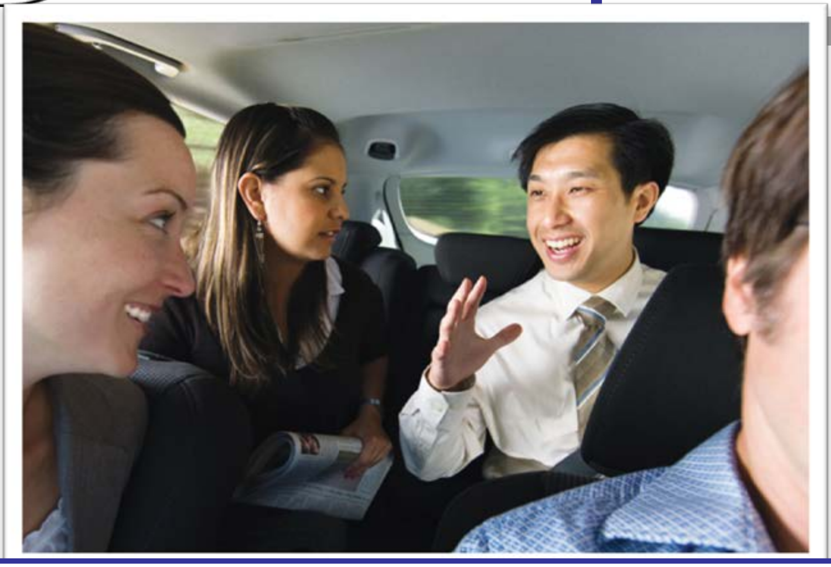
The Smart Society vision

- Computers are great at storing, processing and communicating data
- People are great at interpreting context (semantics), interpersonal relationships and social norms
- Can we combine the best of both worlds to build a “smarter society” via hybrid social computation decentralized through society?

An **holistic** approach, with the goal of building hybrid social systems, capable of handling, exploiting and ***composing, as social computations,*** human and machine ***diverse*** actions, towards a smarter society

www.smart-society-project.eu,
Source: Fausto Giunchiglia

An example: the Ride Sharing



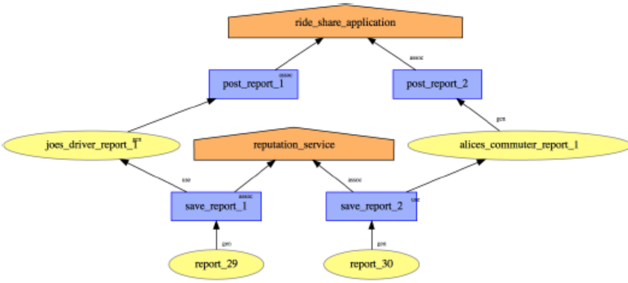
Desired properties

- Recommendation system
- Different profiles, dynamically changing
- Privacy,
- Provenance, trust and reputation
- Sensing and context recognition
- Effective communication among agents
- Coordinate actors
- …

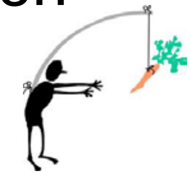
[…] With economic pressures pushing travelers to share resources whenever possible (and various online services making that ever easier), I wondered how this was possible. Did I do something wrong? Were the sites inadequate? Or were there simply not enough people out there looking for such a ride? The answer turned out to be a bit of each. Ride-sharing's move from campus bulletin board to Internet had been less successful than expected ***[New York Times – September 17, 2013]***

www.smart-society-project.eu,
Source: Fausto Giunchiglia

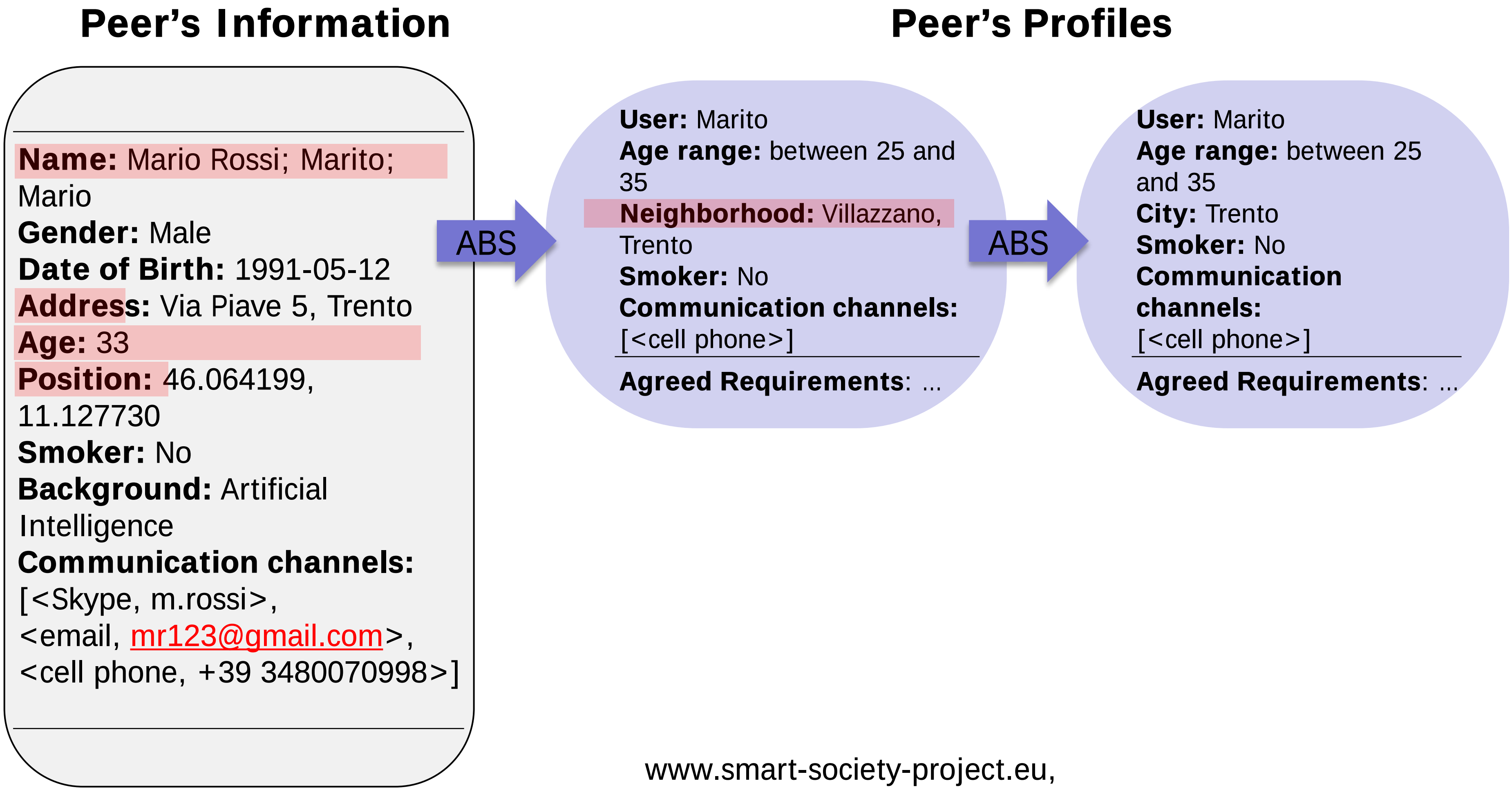
Privacy Challenges



- Peer Profiling and Search
- Provenance, Trust and Reputation
- Incentives and decision making
- Sensed Data
- Entangled Data – who has control?
- Privacy of Collectives



Privacy by information abstraction and sticky policies



Peer's Information

Name: Mario Rossi; Marito; Mario
Gender: Male
Date of Birth: 1991-05-12
Address: Via Piave 5, Trento
Age: 33
Position: 46.064199, 11.127730
Smoker: No
Background: Artificial Intelligence
Communication channels: [<Skype, m.rossi>, <email, mr123@gmail.com>, <cell phone, +39 3480070998>]

Peer's Profiles

User: Marito
Age range: between 25 and 35
Neighborhood: Villazzano, Trento
Smoker: No
Communication channels: [<cell phone>]

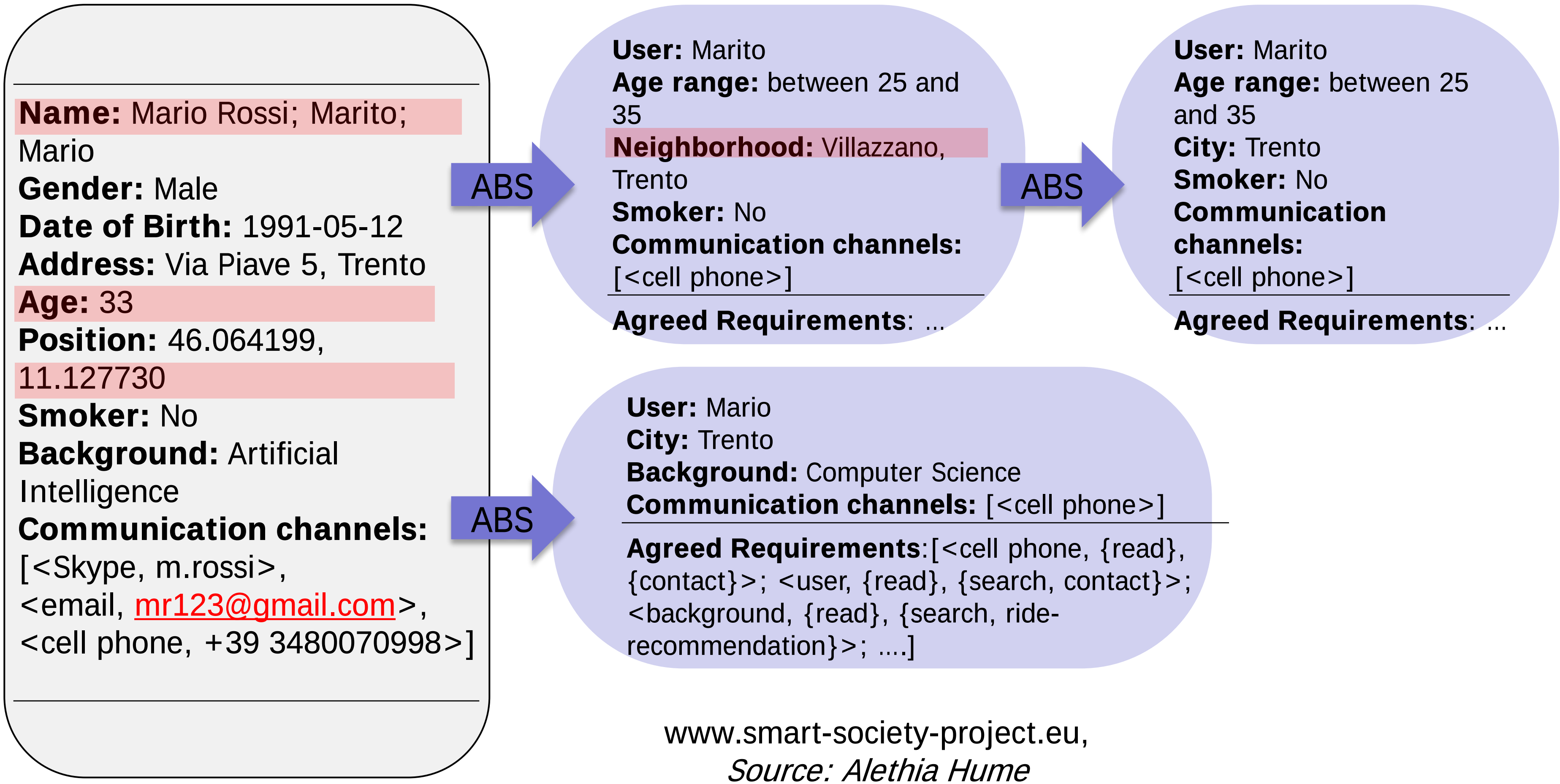
Agreed Requirements: ...

User: Marito
Age range: between 25 and 35
City: Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

www.smart-society-project.eu,
Source: Alethia Hume

Privacy by information abstraction and sticky policies



Peer's Information

Name: Mario Rossi; Marito; Mario
Gender: Male
Date of Birth: 1991-05-12
Address: Via Piave 5, Trento
Age: 33
Position: 46.064199, 11.127730
Smoker: No
Background: Artificial Intelligence
Communication channels: [<Skype, m.rossi>, <email, mr123@gmail.com>, <cell phone, +39 3480070998>]

Peer's Profiles

User: Marito
Age range: between 25 and 35
Neighborhood: Villazzano, Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

User: Marito
Age range: between 25 and 35
City: Trento
Smoker: No
Communication channels: [<cell phone>]

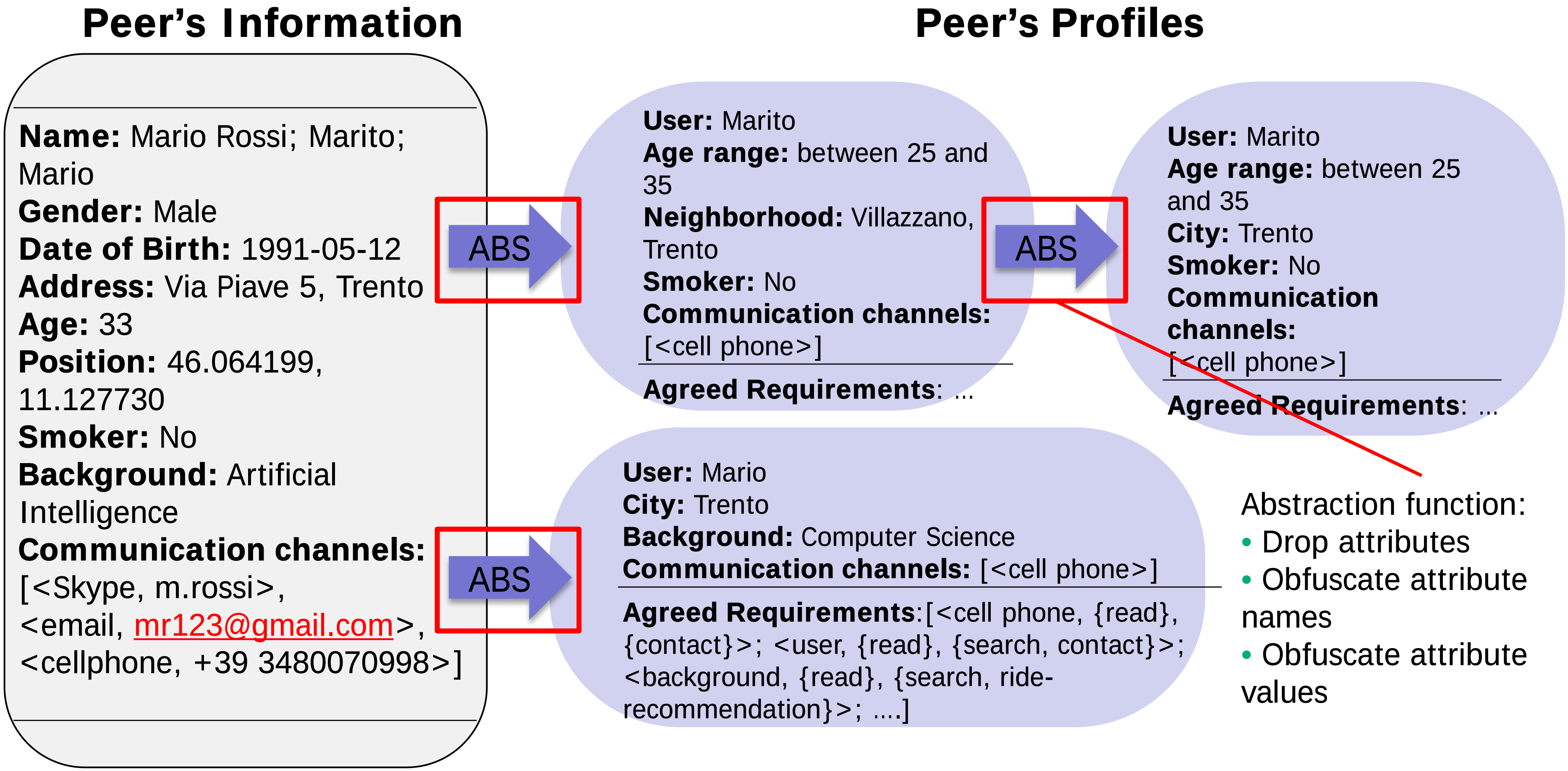
Agreed Requirements: ...

User: Mario
City: Trento
Background: Computer Science
Communication channels: [<cell phone>]

Agreed Requirements:[<cell phone, {read}, {contact}>; <user, {read}, {search, contact}>; <background, {read}, {search, ride-recommendation}>; ....]

www.smart-society-project.eu,
Source: Alethia Hume

Privacy by information abstraction and sticky policies



Peer's Information

Name: Mario Rossi; Marito; Mario
Gender: Male
Date of Birth: 1991-05-12
Address: Via Piave 5, Trento
Age: 33
Position: 46.064199, 11.127730
Smoker: No
Background: Artificial Intelligence
Communication channels: [<Skype, m.rossi>, <email, mr123@gmail.com>, <cellphone, +39 3480070998>]

Peer's Profiles

User: Marito
Age range: between 25 and 35
Neighborhood: Villazzano, Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

User: Marito
Age range: between 25 and 35
City: Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

User: Mario
City: Trento
Background: Computer Science
Communication channels: [<cell phone>]

Agreed Requirements:[<cell phone, {read}, {contact}>; <user, {read}, {search, contact}>; <background, {read}, {search, ride-recommendation}>; ....]

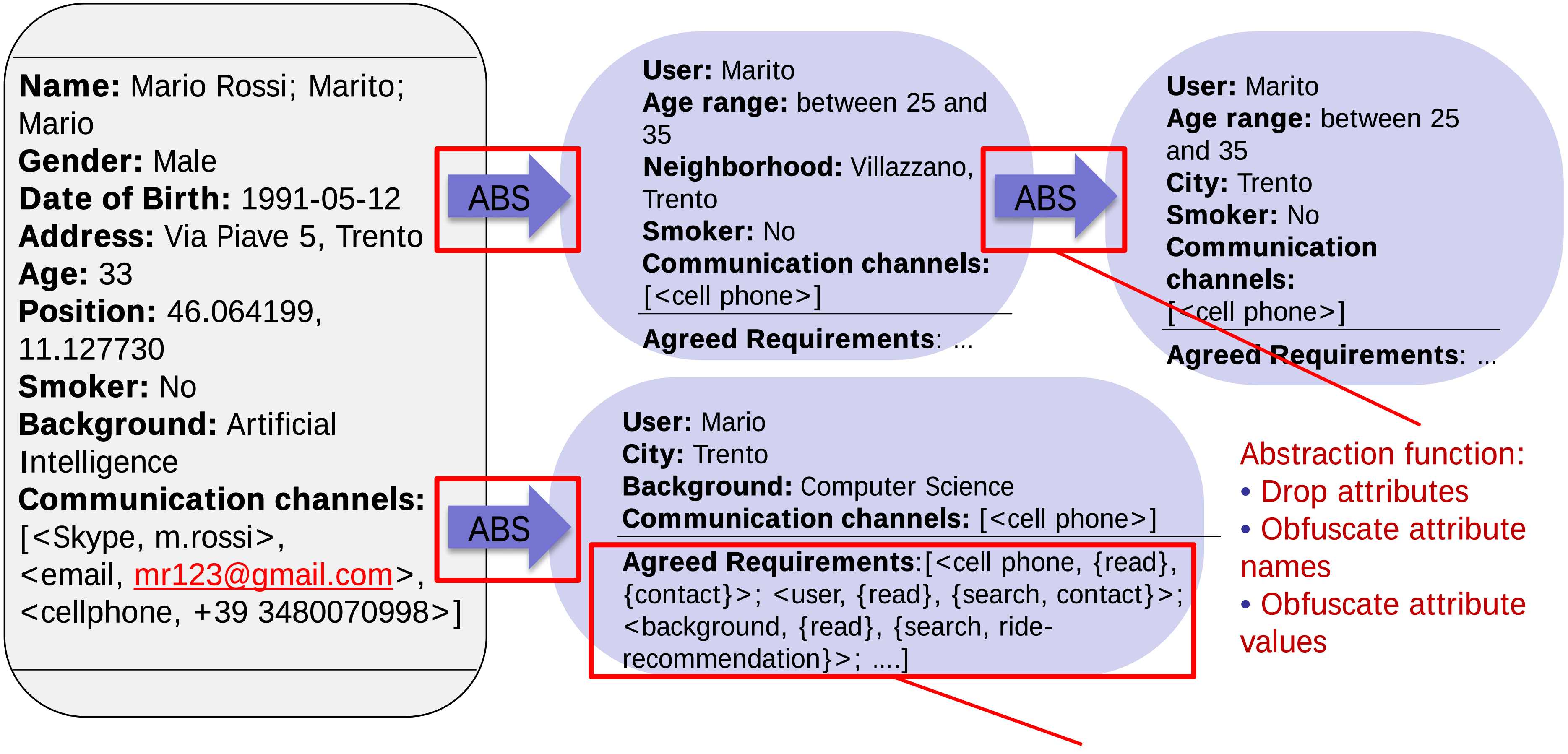
Abstraction function:

- Drop attributes
- Obfuscate attribute names
- Obfuscate attribute values

www.smart-society-project.eu,
Source: Alethia Hume

Privacy by information abstraction and sticky policies

Peer's Information **Peer's Profiles**



Name: Mario Rossi; Marito; Mario
Gender: Male
Date of Birth: 1991-05-12
Address: Via Piave 5, Trento
Age: 33
Position: 46.064199, 11.127730
Smoker: No
Background: Artificial Intelligence
Communication channels: [<Skype, m.rossi>, <email, mr123@gmail.com>, <cellphone, +39 3480070998>]

User: Marito
Age range: between 25 and 35
Neighborhood: Villazzano, Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

User: Marito
Age range: between 25 and 35
City: Trento
Smoker: No
Communication channels: [<cell phone>]

Agreed Requirements: ...

User: Mario
City: Trento
Background: Computer Science
Communication channels: [<cell phone>]

Agreed Requirements:[<cell phone, {read}, {contact}>; <user, {read}, {search, contact}>; <background, {read}, {search, ride-recommendation}>; ....]

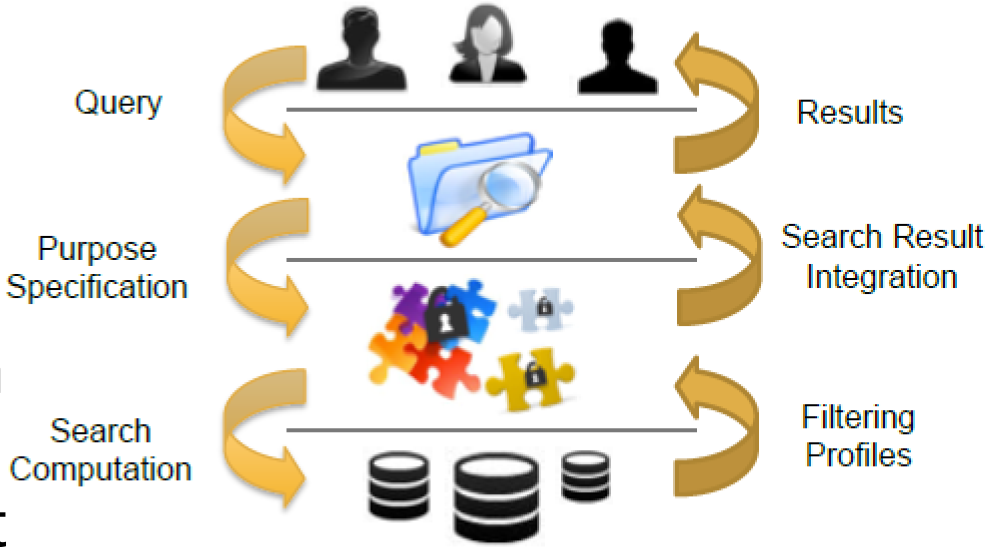
Abstraction function:

- Drop attributes
- Obfuscate attribute names
- Obfuscate attribute values

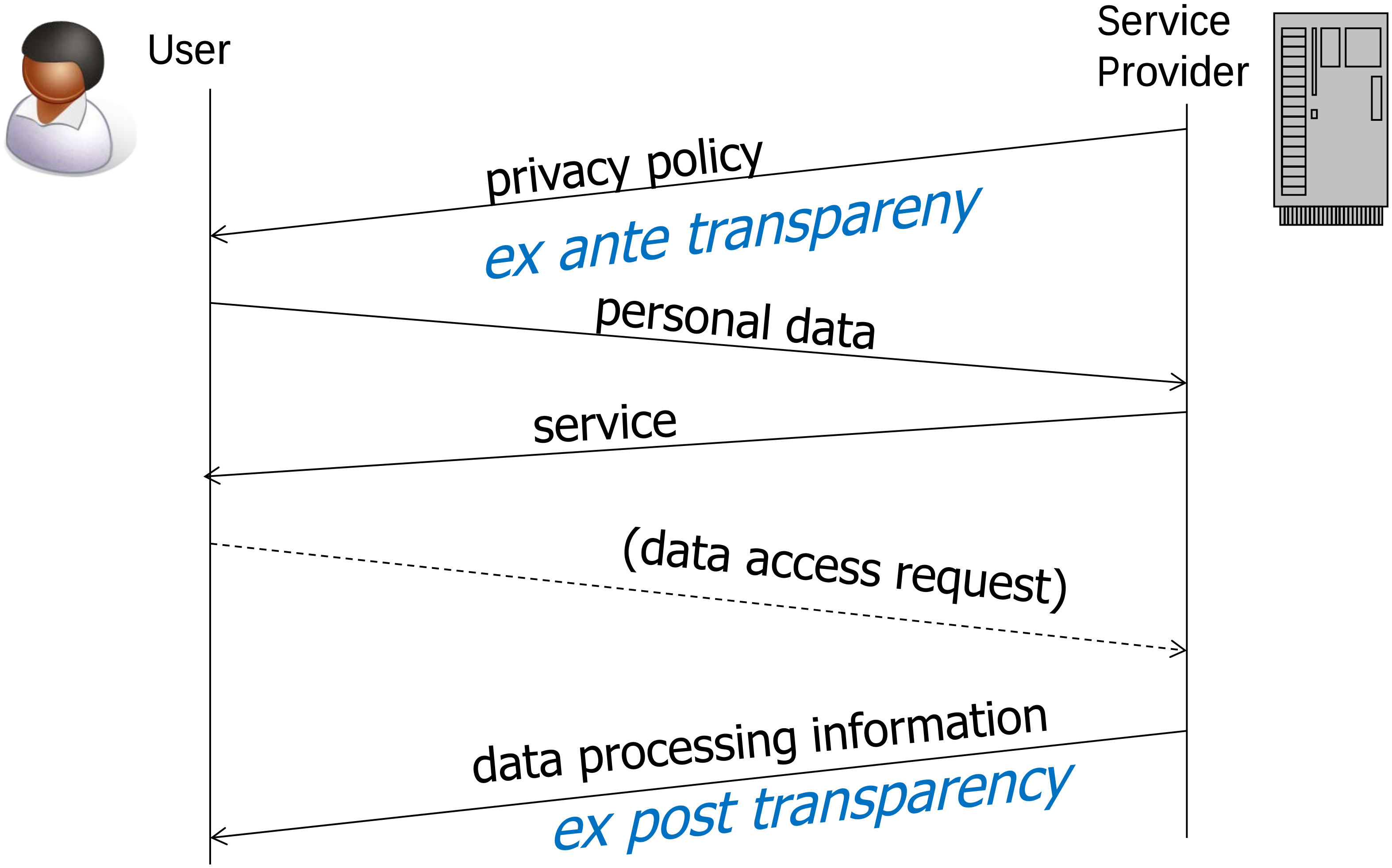
Sticky policies using A-PPL

Privacy-aware Search

- Data minimisation: Search is performed on peer's profiles (i.e. on partial or obfuscated data)
- Purpose binding: Decision on what profile information can be included in the search for what purposes it can be used is regulated by sticky policies



III. Transparency Enhancing Tools (TETs)



User

Service Provider

privacy policy

ex ante transpareny

personal data

service

(data access request)

data processing information

ex post transparency

Transparency & Intervenability

- Legal privacy principles
  - *EU Data Protection Directive*:
    - Informed consent
    - Rights to information, notification
    - Right to access to data & logics involved in automatic processing (algorithm, decision criteria, source of data)
    - Rights to correction/blocking/deletion
    - Right not to be subject to automated decisions
  - *Proposed EU General Data Protection Regulation (GDPR)*
    - Right to object to profiling, data breach notification, exercising data subject rights electronically
  - *Swedish Data Patient Act*:
    - Rights to access health records and log information
- Social Trust Factors
  - Increased trust in applications if procedures are clear, transparent and reversible

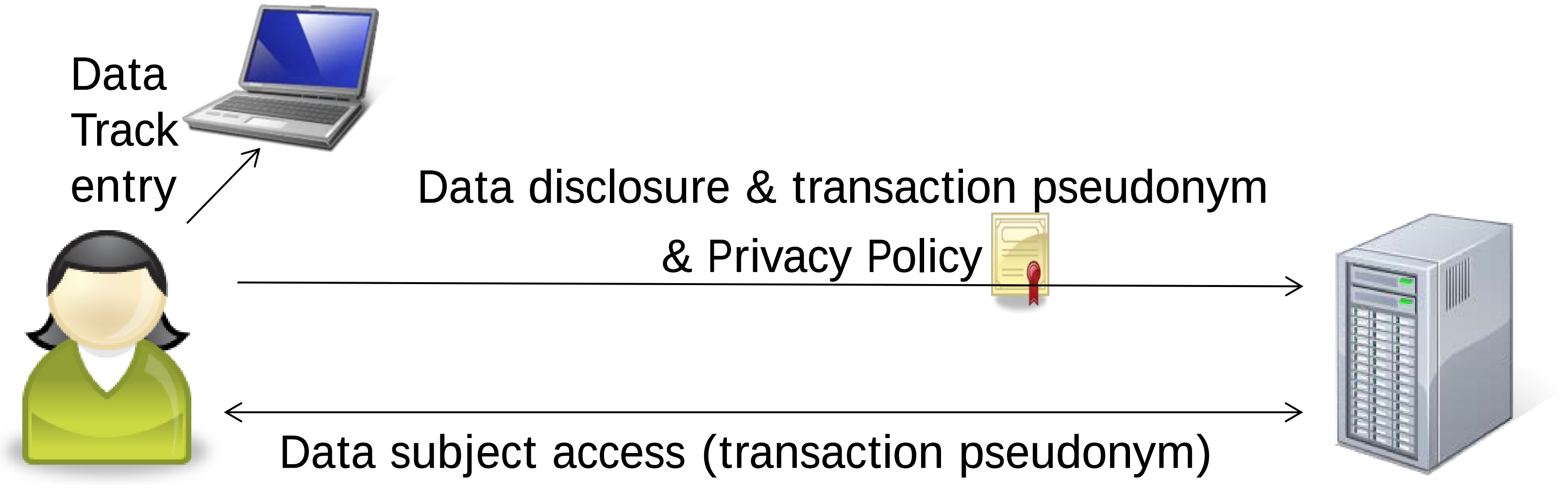
Transparency vs. Confidentiality
Examples

- **Log files in eHealth – privacy issues:**
  - Information about who (e.g., psychi accessed EHR is sensitive for p
  - Monitoring of performan medical personnel

- **Business on to profi**
  - (cf. R rotection Directive)

Requirements:
- Privacy-preserving
- Tradeoff with Business Secrets

Ex post TETs: Data Track



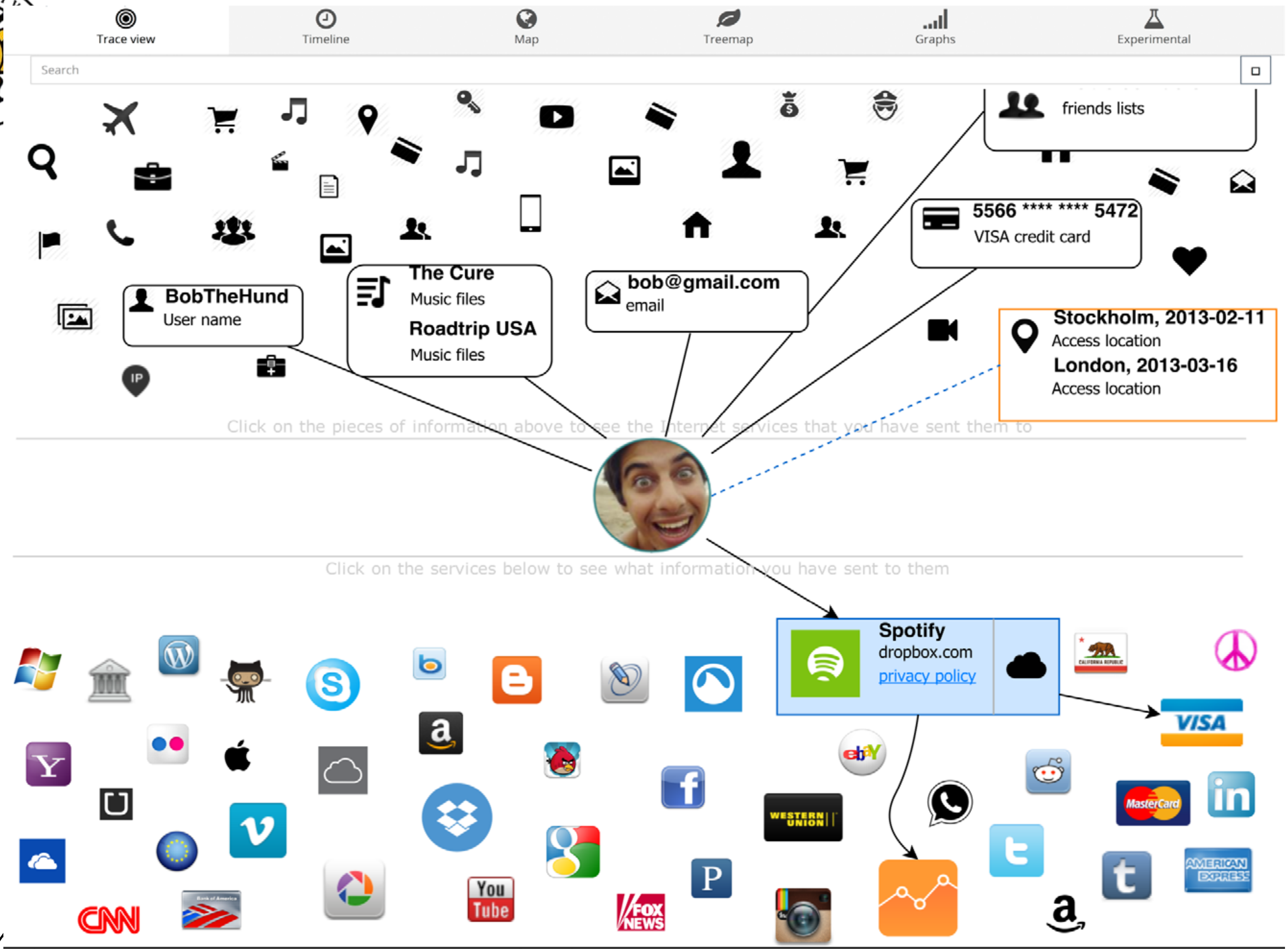
Data Track entry

Data disclosure & transaction pseudonym & Privacy Policy

Data subject access (transaction pseudonym)

CLOUD ACCOUNTABILITY PROJECT

A4Cloud Data Track (3rd Iteration)



friends lists

5566 **** **** 5472
VISA credit card

BobTheHund
User name

The Cure
Music files

Roadtrip USA
Music files

bob@gmail.com
email

Stockholm, 2013-02-11
Access location

London, 2013-03-16
Access location

Click on the pieces of information above to see the Internet services that you have sent them to

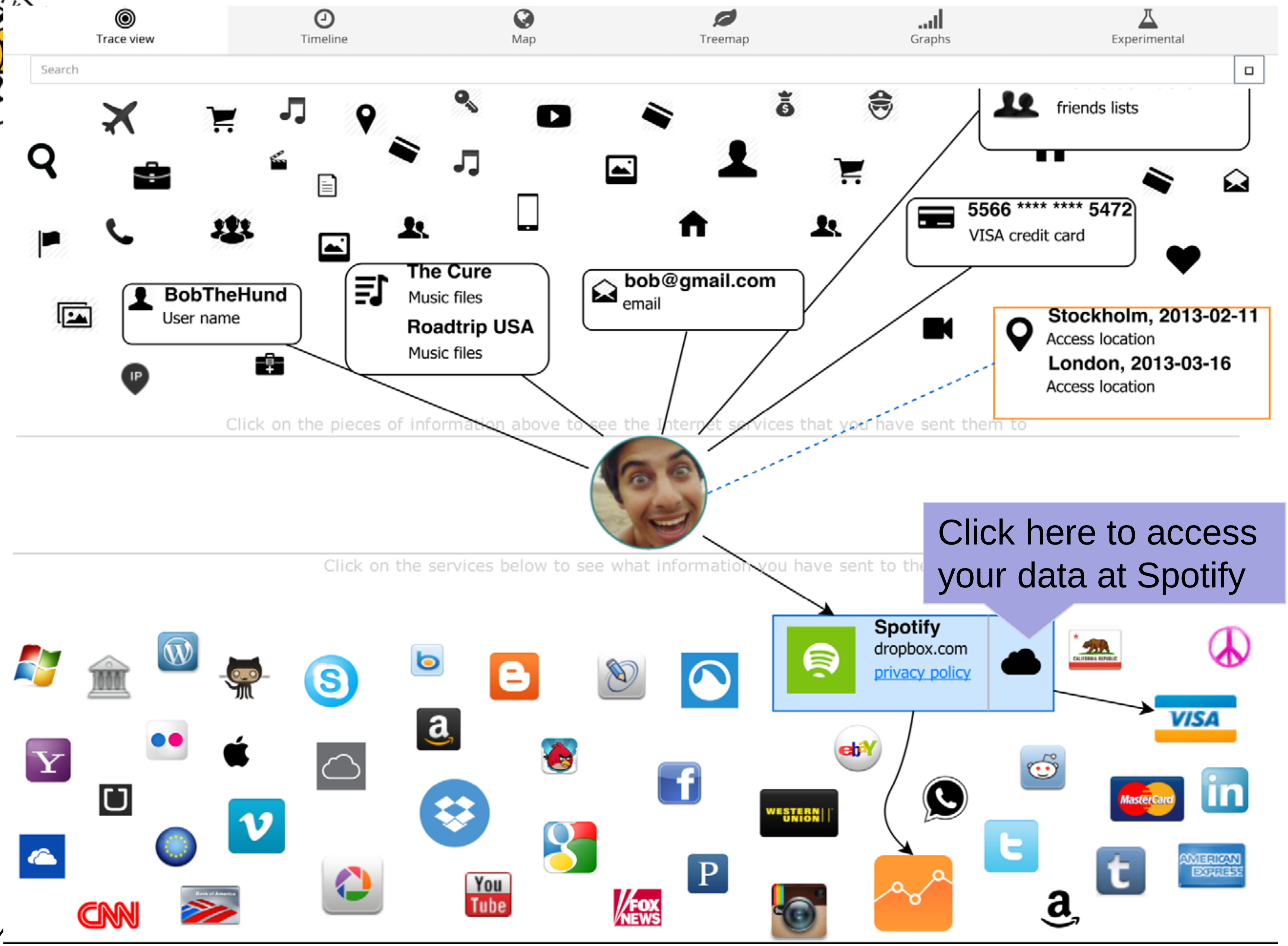
Click on the services below to see what information you have sent to them

Spotify
dropbox.com
privacy policy

Source: /

A4Cloud Data Track

CLOUD ACCOUNTABILITY PROJECT



Trace view | Timeline | Map | Treemap | Graphs | Experimental

Search

friends lists

5566 **** **** 5472
VISA credit card

BobTheHund
User name

The Cure
Music files

Roadtrip USA
Music files

bob@gmail.com
email

Stockholm, 2013-02-11
Access location

London, 2013-03-16
Access location

Click on the pieces of information above to see the Internet services that you have sent them to

Click on the services below to see what information you have sent to the

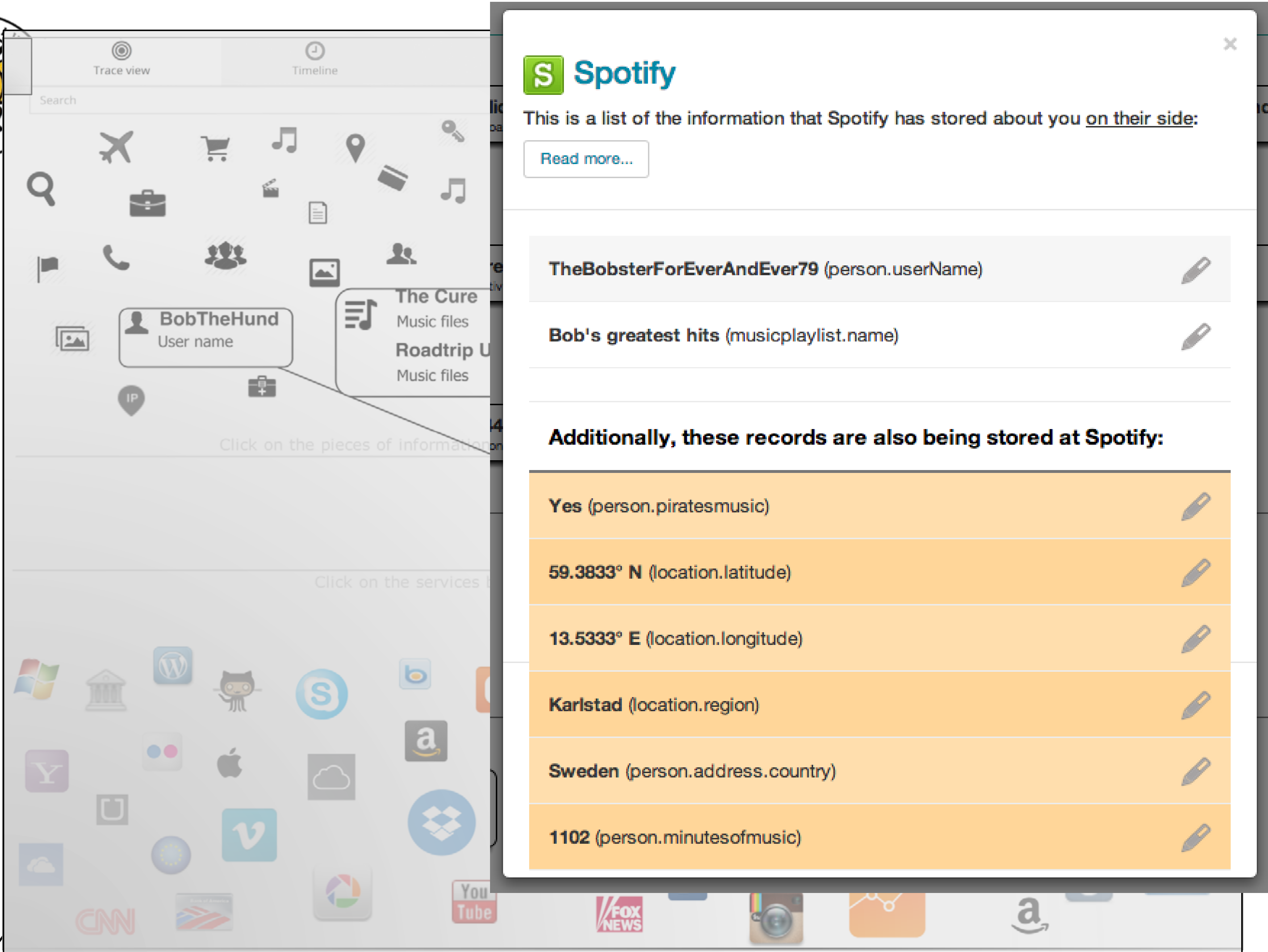
Spotify
dropbox.com
privacy policy

Click here to access your data at Spotify

Source: /

A4Cloud Data Track

CLOUD ACCOUNTABILITY PROJECT



Spotify

This is a list of the information that Spotify has stored about you on their side:

Read more...

TheBobsterForEverAndEver79 (person.userName)

Bob's greatest hits (musicplaylist.name)

Additionally, these records are also being stored at Spotify:

Yes (person.piratesmusic)

59.3833° N (location.latitude)

13.5333° E (location.longitude)

Karlstad (location.region)

Sweden (person.address.country)

1102 (person.minutesofmusic)

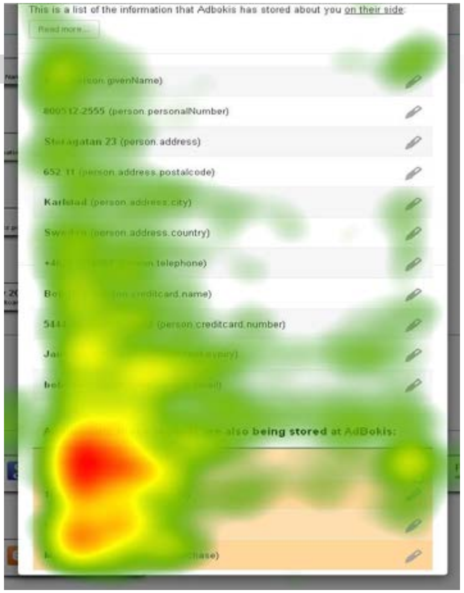
Source: /

User evaluations

- Do users find the trace view of the Data Track **intuitive and comprehensible**?
- Do users appreciate the **functionality** of the Data Track?
- Do users understand that there are **two different views** (data records stored under the users' control (locally or in a privacy-friendly cloud infrastructure) and data records stored at the service provider)

Usability test results

Where are the Data Track records stored?	Frequency	Percent
On the DataTrack program (on a cloud/Internet storage)	9	52.9
On the DataTrack program (locally in computer)	4	23.5
On the Internet somewhere	1	5.9
On the services that I have given information to	3	17.6



This is a list of the information that Adbokis has stored about you on their side:

Read more...

...son.givenName)

800512-2555 (person.personalNumber)

Storagatan 23 (person.address)

652 11 (person.address.postalcode)

Karlstad (person.address.city)

Sweden (person.address.country)

+4... (person.telephone)

Bo... (person.creditcard.name)

544... (person.creditcard.number)

Ja...

bo...

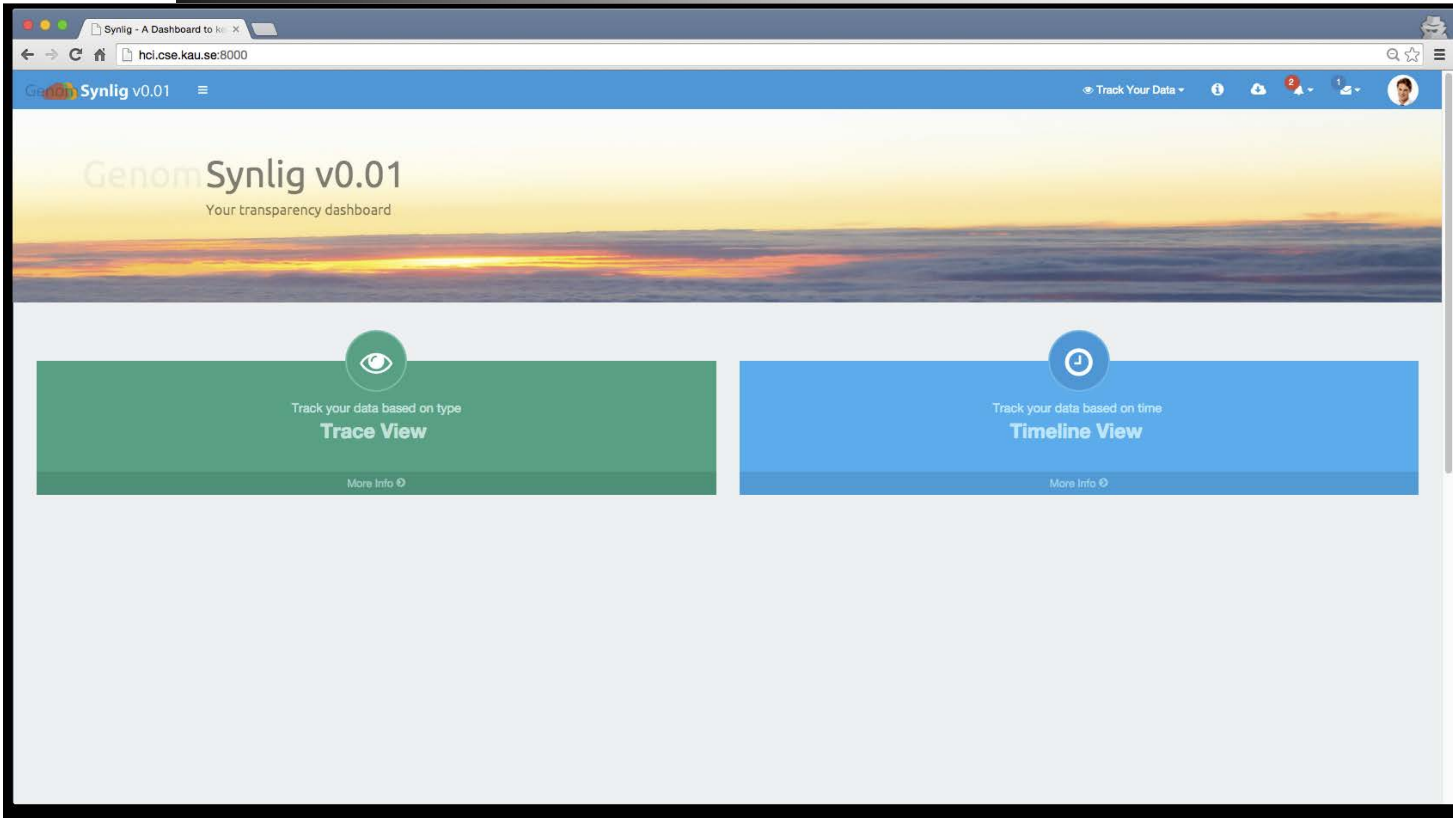
...also being stored at AdBokis:

...chase)

How often do you think you would have the program turned on so that it tracks the information you give to Internet services?	
Never tracking (-0% of the time)	2
Rarely tracking (25% of the time)	3
Often tracking (75% of the time)	5
Always tracking (100% of the time)	7

How often do you believe you would use the DataTrack program?	
Very rarely (almost never or never)	1
Rarely (a few times per year)	1
Sometimes (a few times per month)	7
Often (around two to four times per week)	4
Very often (almost always)	4

4th Iteration – Data Track *”GenomSynlig” Dashboard*



GenomSynlig v0.01

Your transparency dashboard

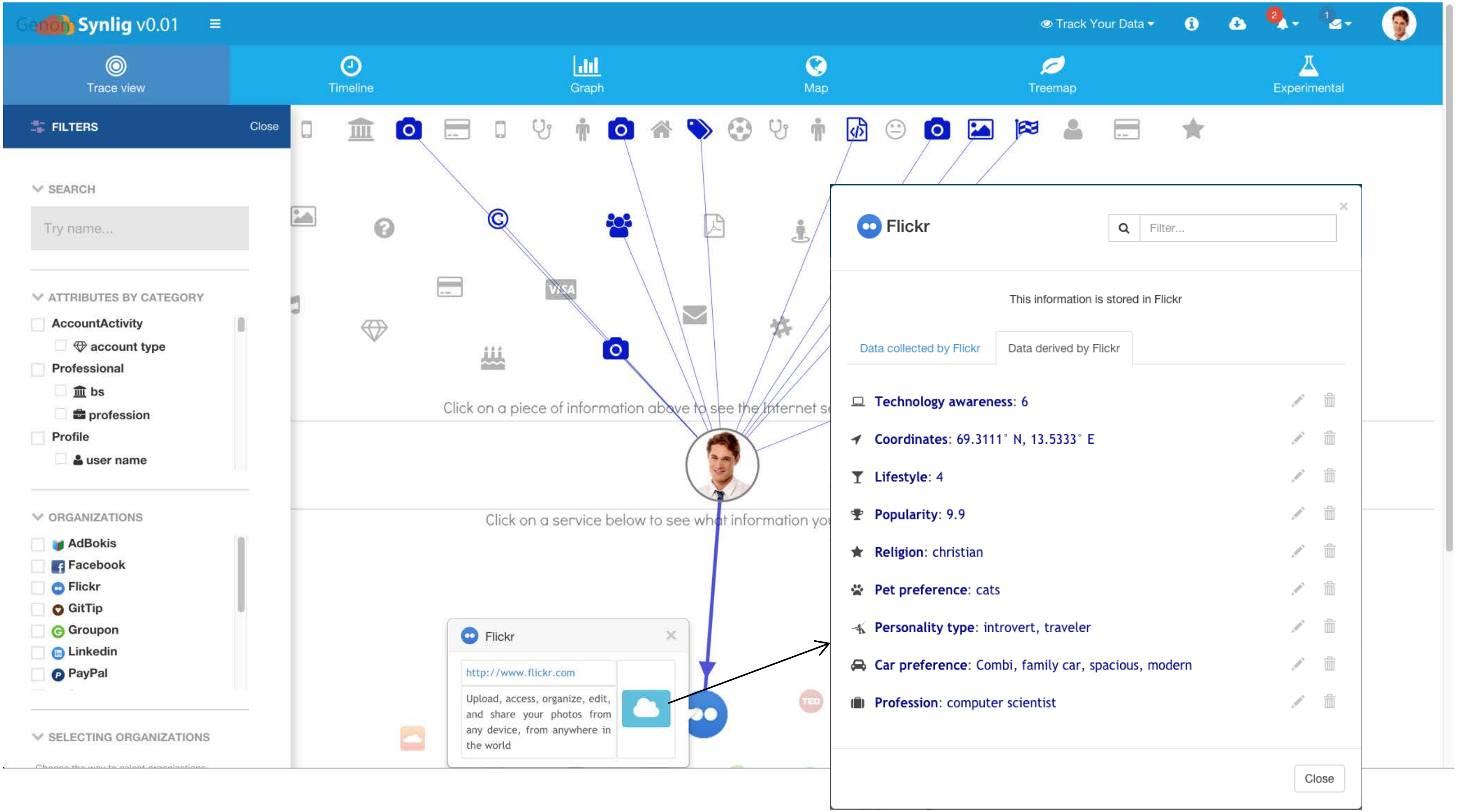
Track your data based on type

Trace View

Track your data based on time

Timeline View

Trace View



ATTRIBUTES BY CATEGORY

- □ AccountActivity
  - □ account type
- □ Professional
  - □ bs
  - □ profession
- □ Profile
  - □ user name

ORGANIZATIONS

- □ AdBokis
- □ Facebook
- □ Flickr
- □ GitTip
- □ Groupon
- □ Linkedin
- □ PayPal

SELECTING ORGANIZATIONS

Click on a piece of information above to see the Internet s

Click on a service below to see what information yo

Flickr

http://www.flickr.com

Upload, access, organize, edit, and share your photos from any device, from anywhere in the world

Flickr

Filter...

This information is stored in Flickr

Data collected by Flickr | Data derived by Flickr

Technology awareness: 6

Coordinates: 69.3111° N, 13.5333° E

Lifestyle: 4

Popularity: 9.9

Religion: christian

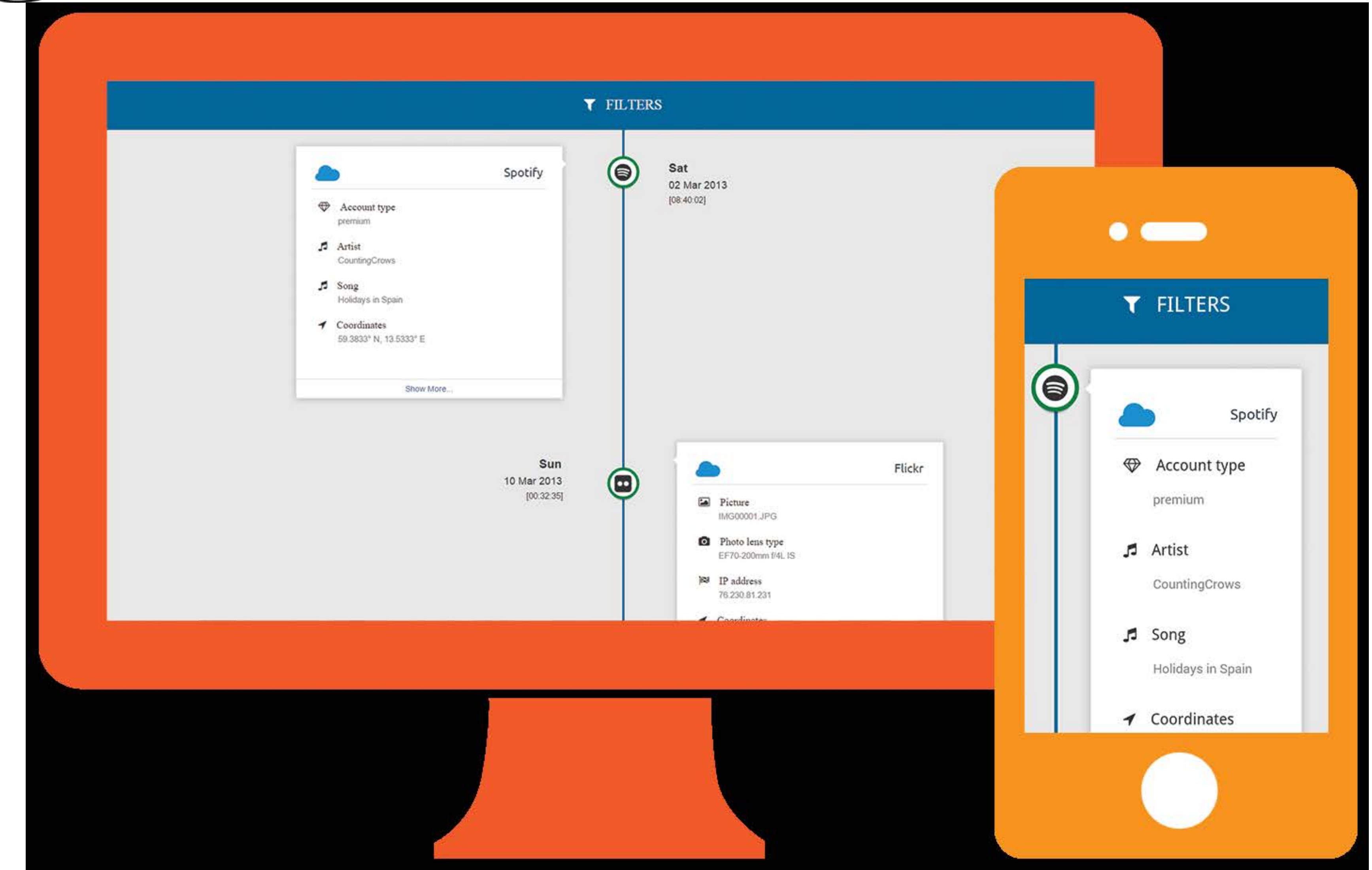
Pet preference: cats

Personality type: introvert, traveler

Car preference: Combi, family car, spacious, modern

Profession: computer scientist

Timeline View



FILTERS

Spotify

Account type
premium

Artist
CountingCrows

Song
Holidays in Spain

Coordinates
59.3833° N, 13.5333° E

Show More...

Sat
02 Mar 2013
[08:40:02]

Sun
10 Mar 2013
[00:32:35]

Flickr

Picture
IMG00001.JPG

Photo lens type
EF70-200mm f/4L IS

IP address
76.230.81.231

Coordinates

FILTERS

Spotify

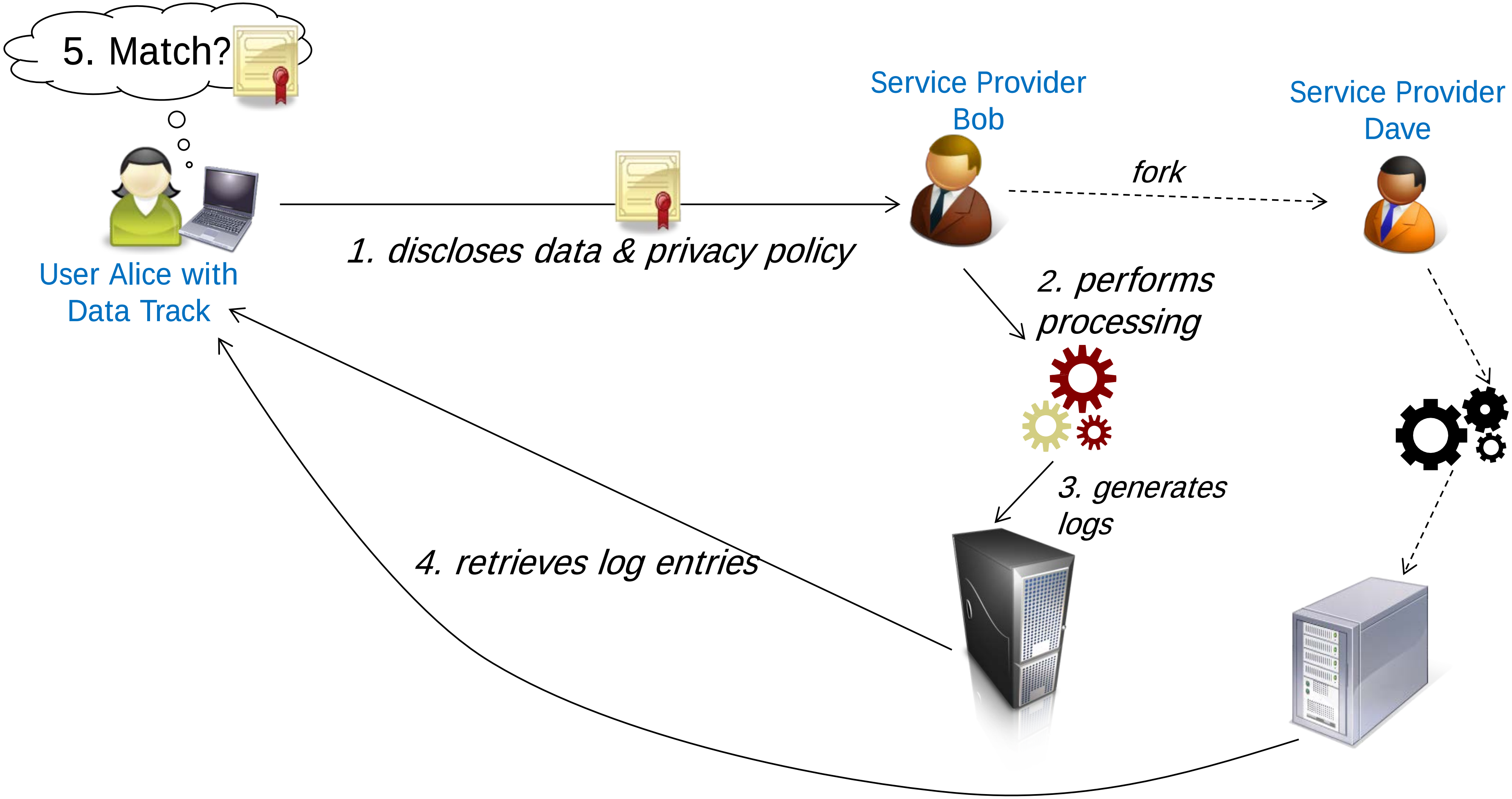
Account type
premium

Artist
CountingCrows

Song
Holidays in Spain

Coordinates

Distributed Transparency Logging



Source: T. Pulls et al., Privacy-Preserving Transparency Logging, WPES 2013

Distributed Privacy-Preserving Transparency Logging – *Security Properties*

- Integrity: no undetectable modifications to logged data (committed prior to compromise)
- Secrecy: only the data subject (or auditor) can read the data logged for him
- Unlinkability of log entries and user identifiers accross data processors

IV. Conclusions

- Privacy principles for Profiling can be enforced by PETs:
  - Anonymisation & Obfuscation
  - Sticky policies for purpose limitation
  - TETs
  - ....
- TETs need to address privacy/Confidentiality requirements
- HCI Requirements for TETs important, e.g.:
  - Make difference between locally and remotely stored data obvious;
  - Provide transparency also for implicitly obtained or derived data;
  - Provide users with help (e.g., via tooltips or introductory tutorials) for understanding how control functions can be activated.

(see also: Angulo, Fischer-Hübner, et.al., CHI 2015 – Proceedings, Work in Progress Session ACM).

Questions ?

http://www.cs.kau.se/~simone/

CLOUD
ACCOUNTABILITY
PROJECT

Smart Society