UNIVERSITY OF CAMBRIDGE

EPSRC
Engineering and Physical Sciences
Research Council

BBC
R&D

Containing Personal Data Processing with the DATABOX

01000100 01100001 01110100 01100001 01100010 01111000

Richard Mortier, SRG, Cambridge University Computer Lab

Hamed Haddadi, EECS, Queen Mary London

horizon
DIGITAL ECONOMY RESEARCH

The University of
Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Queen Mary
University of London

Networks & Operating Systems
SRG, Computer Laboratory

Living in a Big Data World

- Challenges vs Opportunities
  - Who’s tracking us, to what end?
  - Personalisation, Internet of Things
- Digital Footprints
  - Intimate information collected
  - Gathered into large, rich data silos
  - Never forgets or forgives

Key Challenge:

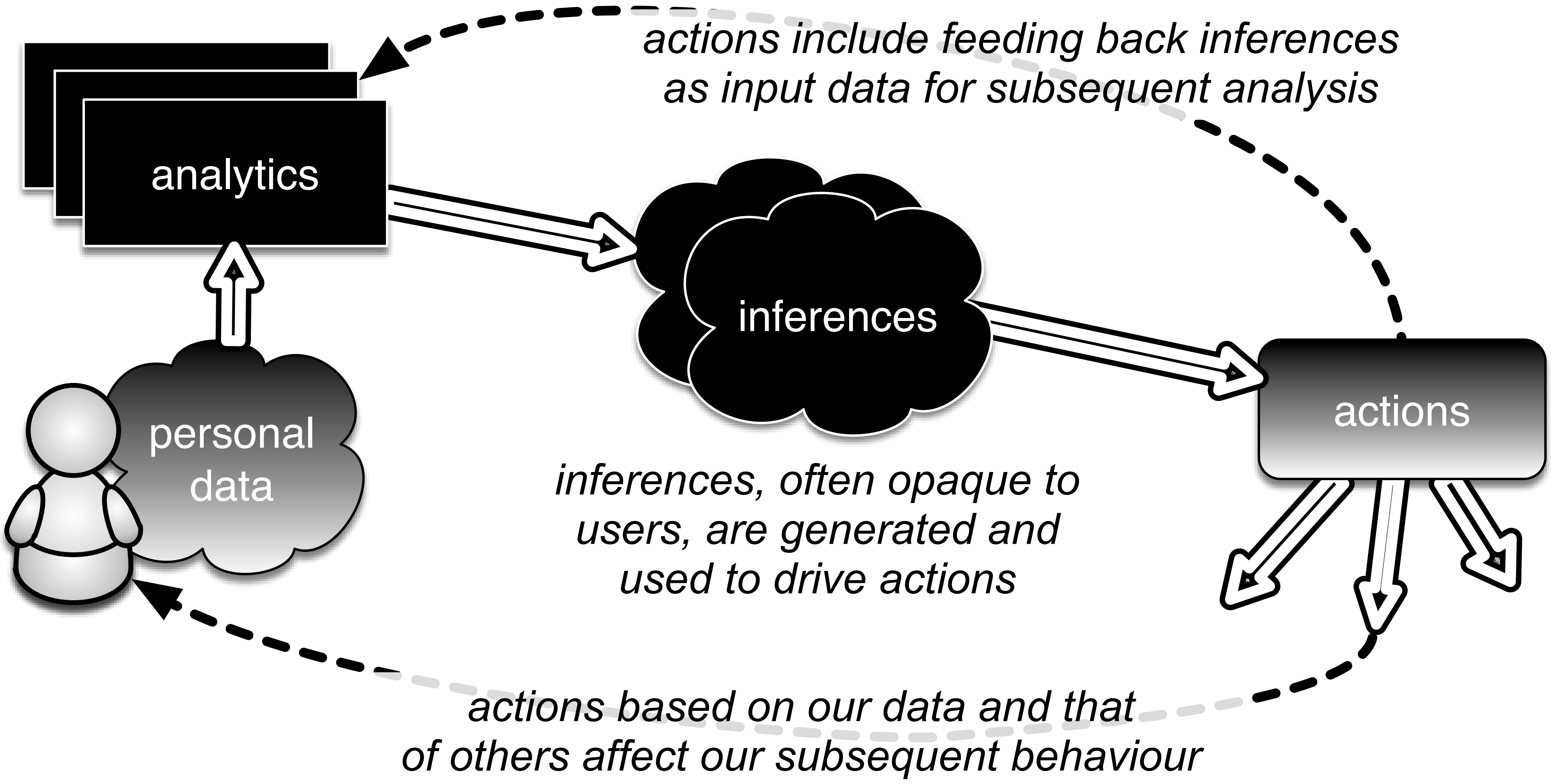
How do we enable data subjects to control collection and exploitation of both **their data** and **data about them**?

http://bigdatapix.tumblr.com/ “Big Data is visualized in so many ways... all of them blue and with numbers and lens flare.”

UNIVERSITY OF CAMBRIDGE

http://weputachipinit.tumblr.com/ “It was just a dumb thing. Then we put a chip in it. Now it's a smart thing.”

Human-Data Interaction



actions include feeding back inferences as input data for subsequent analysis

inferences, often opaque to users, are generated and used to drive actions

actions based on our data and that of others affect our subsequent behaviour

We believe current systems lack

Legibility, **Agency**, **Negotiability**

An Underlying Structural Problem

- The Internet is fragmented, distributed systems are difficult
  - Centralising simplifies things
  - With the cloud, we can, so we do!

- Ease of cloud computing has led to two suboptimal defaults:
  1. Move the data … (by copying)
  2. … to a centralised location

https://www.stickermule.com/marketplace/3442-there-is-no-cloud

Implications

http://cliparts.co/honey-pot-clip-art

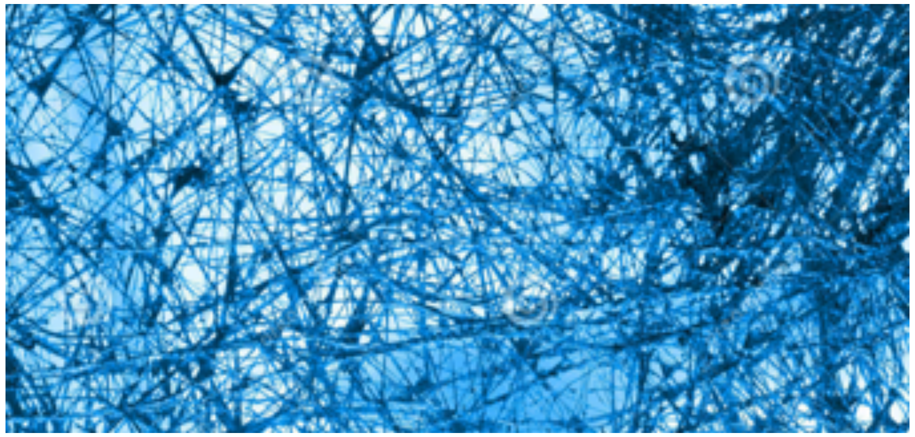
Resilience

- Creation of a honey-pot
- Hidden dependencies

Performance

- Creation of a performance challenge
- Require enormous, reliable, connected resource

http://autoguide.com.vsassets.com/blog/wp-content/uploads/2014/05/traffic-jam.jpg

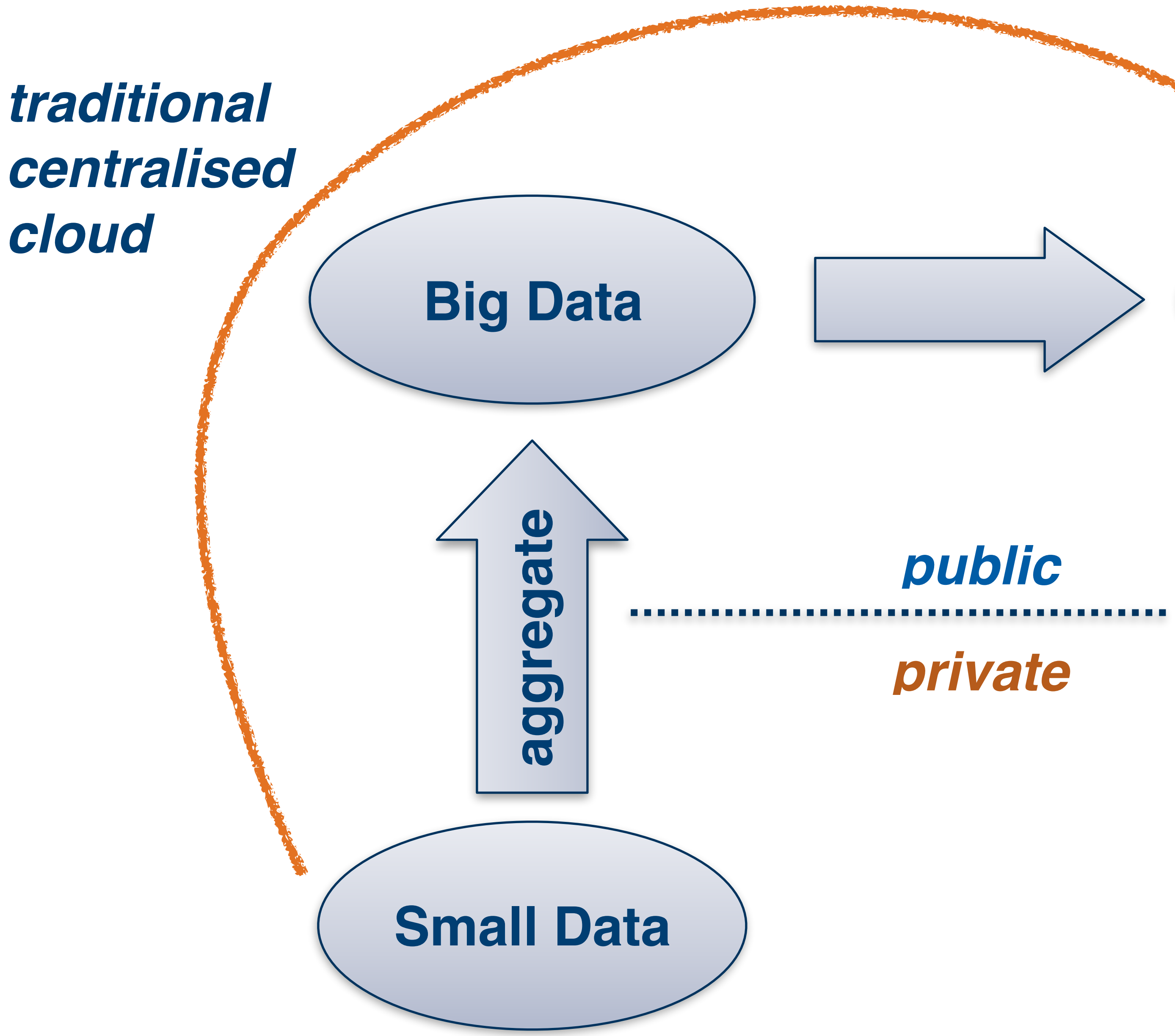


https://www.dreamstime.com/royalty-free-stock-photography-complex-abstract-communication-image18615337

Interaction

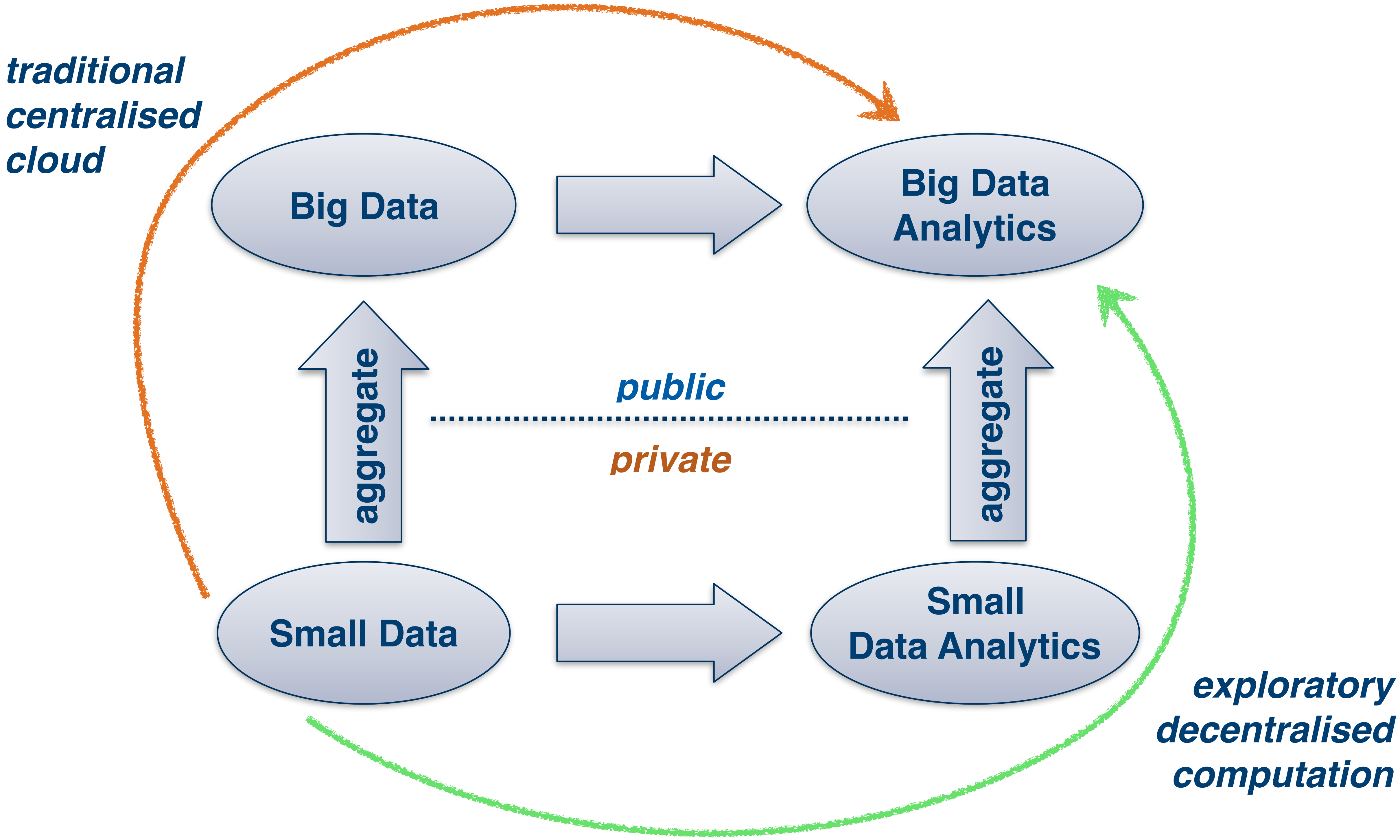
- Abstract “it’s out there somewhere”
- What happens when the Internet goes down?

Big Data Analytics?



- Loss of contextual information
- Ethical and legal issues arise
- Platform technology challenges

Big Data Analytics? Small Data Analytics!



traditional centralised cloud

Big Data

Big Data Analytics

aggregate

public

private

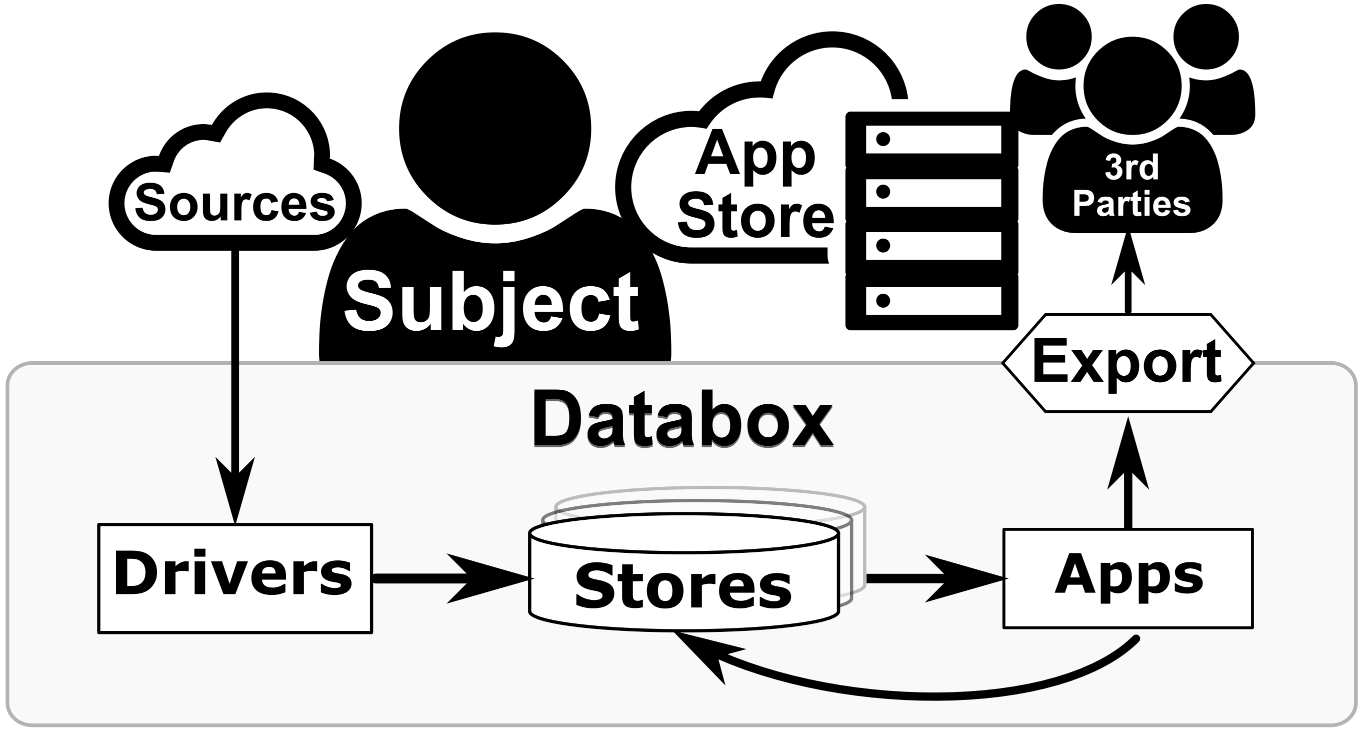
aggregate

Small Data

Small Data Analytics

exploratory decentralised computation

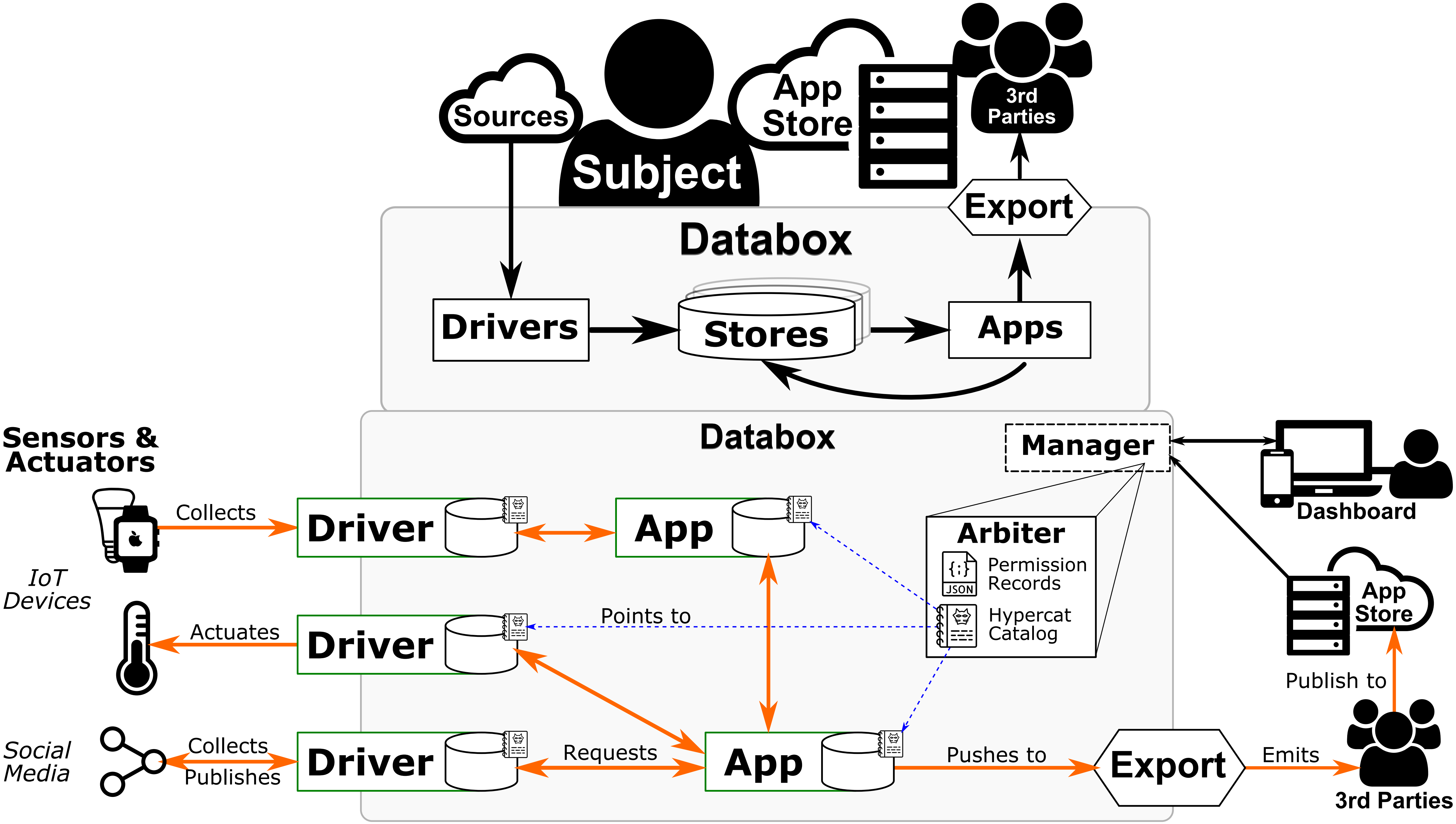
Databox



- Mediates access to data, stored locally as appropriate
- Computations (*apps*) move to data, not data to compute
- Maintain control over internal comms and export
- All operations logged for users to inspect, control

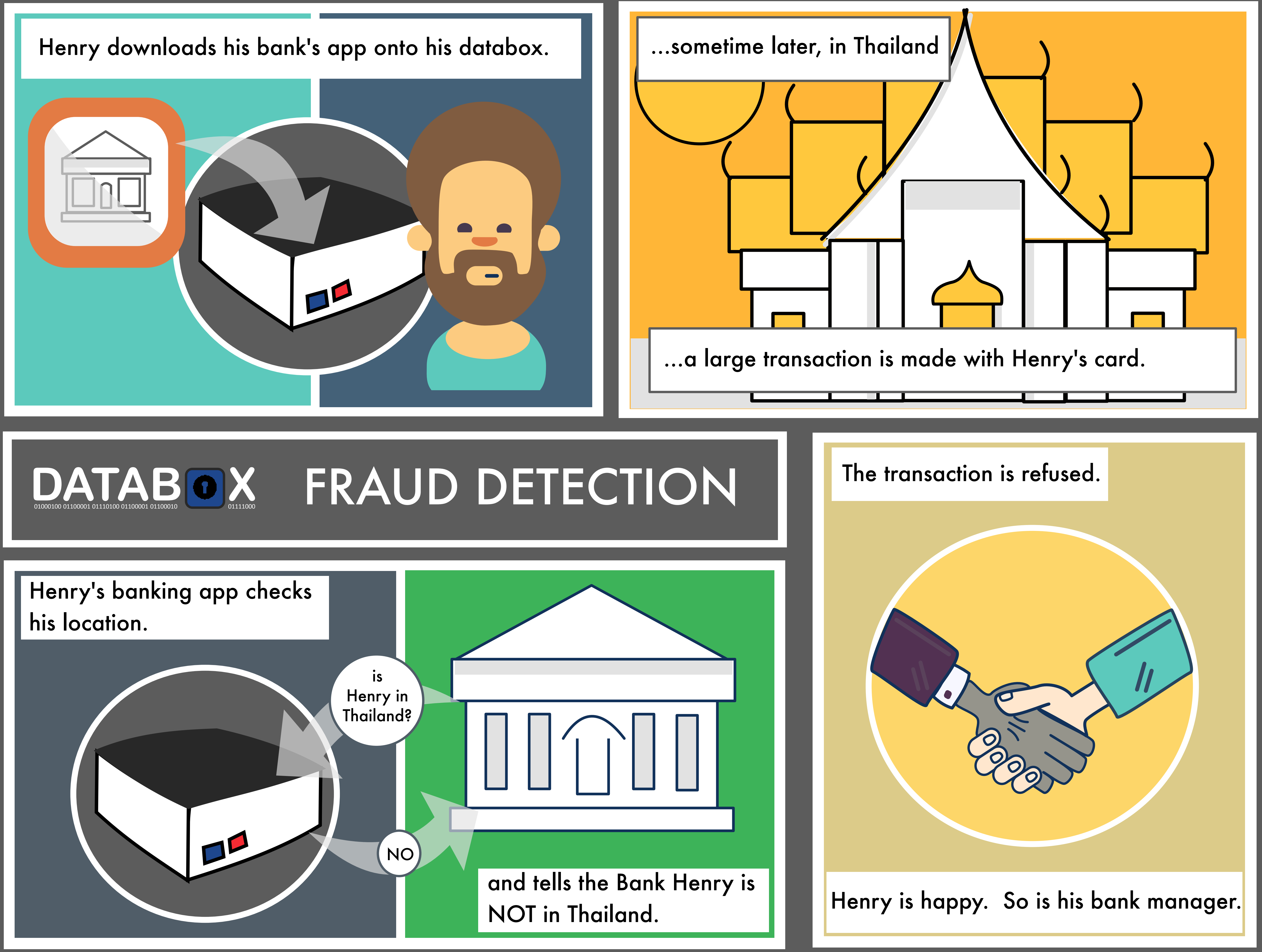
Databox Platform

In submission to SOSP 2017



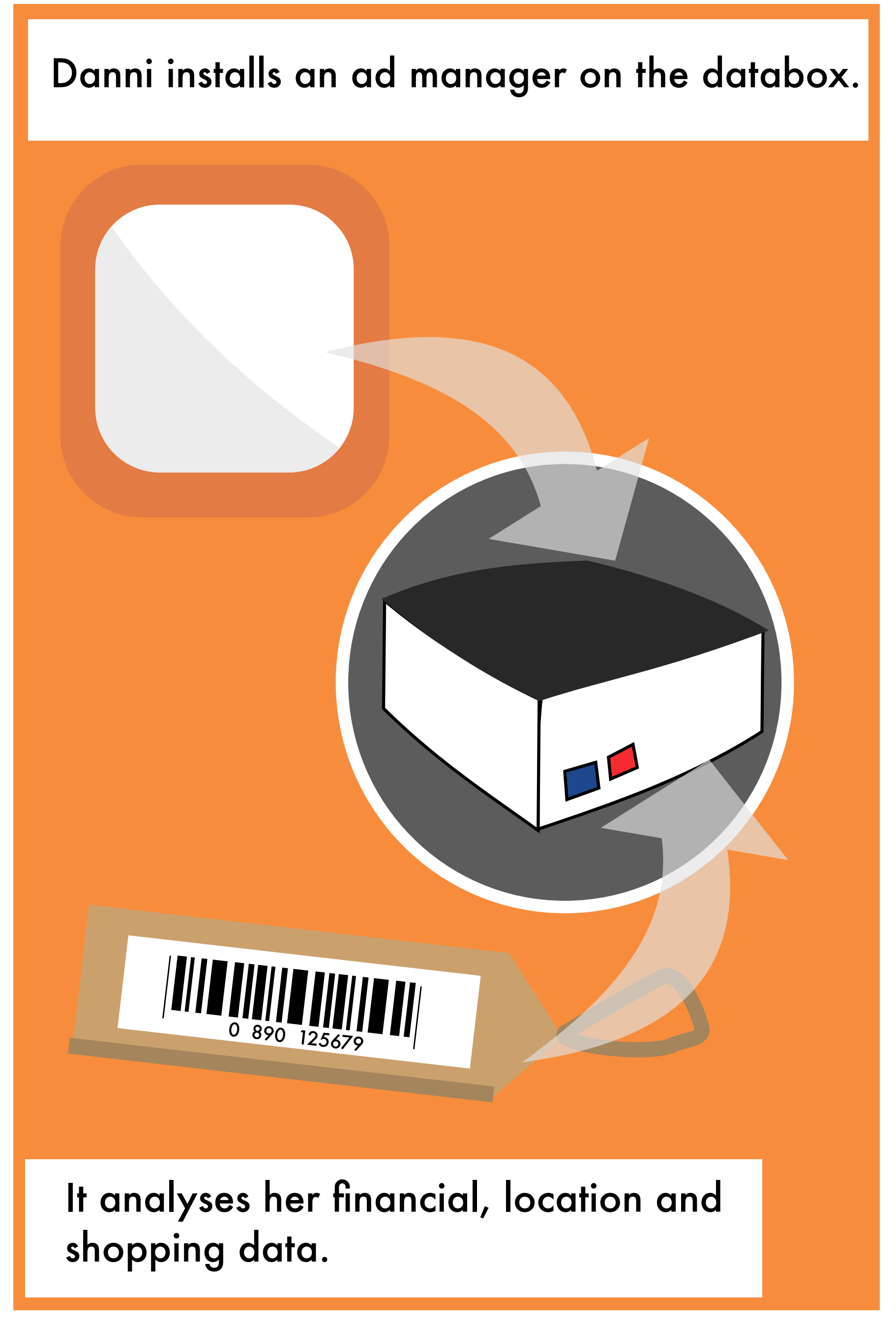
https://github.com/me-box/databox

DATABOX FRAUD DETECTION


Henry downloads his bank's app onto his databox.
...sometime later, in Thailand
...a large transaction is made with Henry's card.
DATABOX
01000100 01100001 01110100 01100001 01100010
01111000
Henry's banking app checks his location.
is Henry in Thailand?
NO
and tells the Bank Henry is NOT in Thailand.
The transaction is refused.
Henry is happy. So is his bank manager.

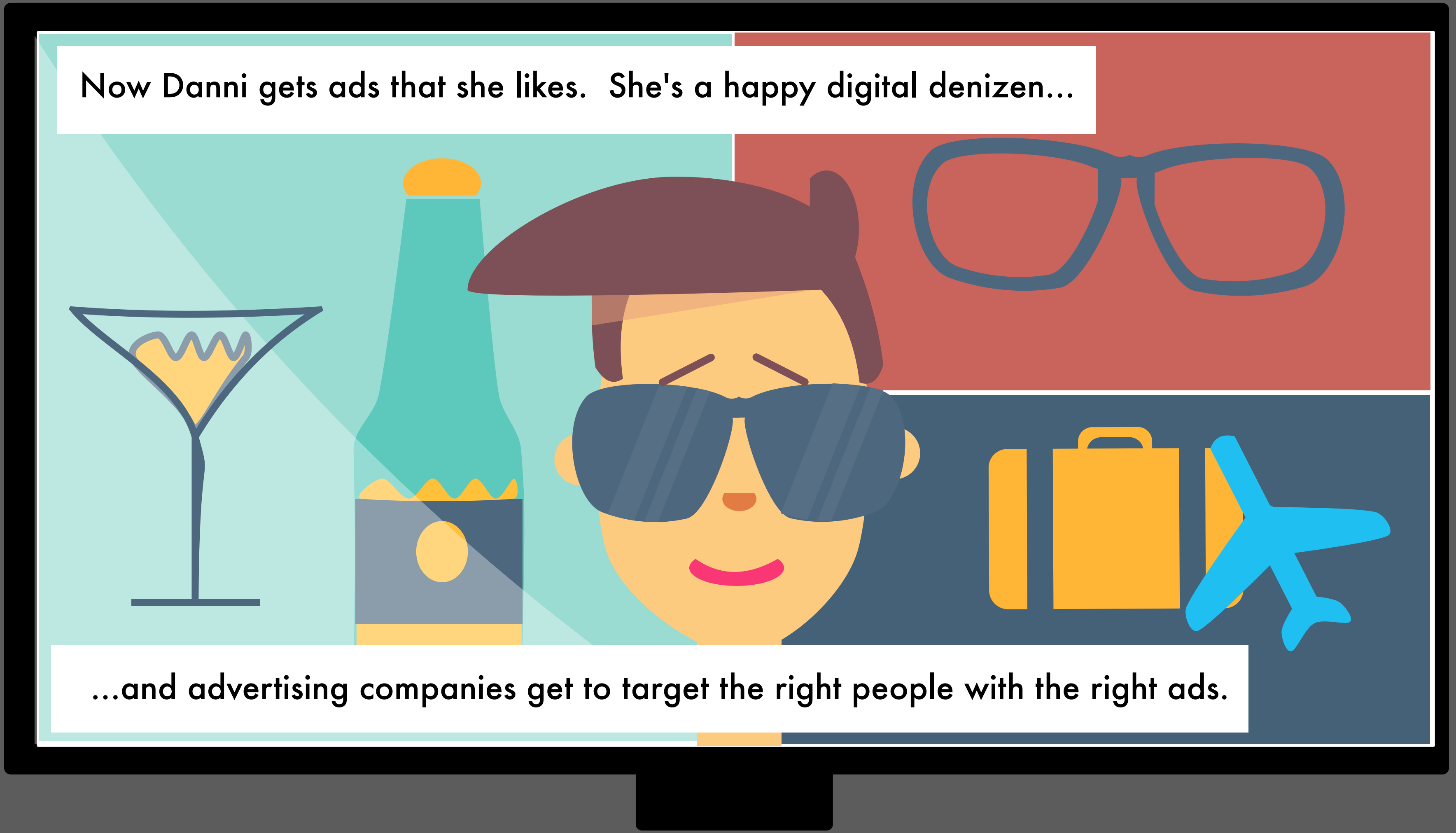
Danni is fed up of being bombarded by random online ads.

She buys a databox to stop them.



Danni installs an ad manager on the databox.

It analyses her financial, location and shopping data.



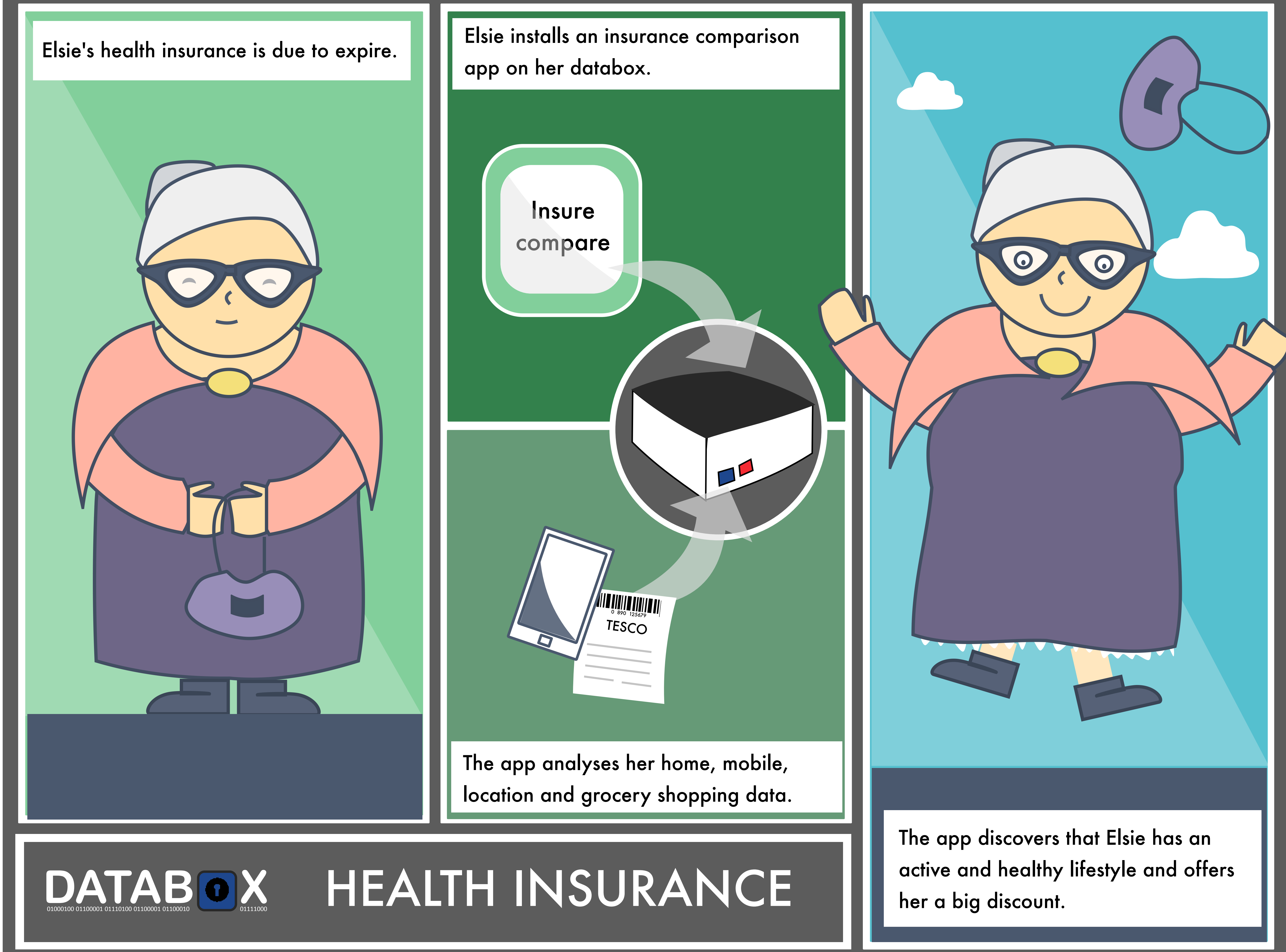
Now Danni gets ads that she likes. She's a happy digital denizen...

...and advertising companies get to target the right people with the right ads.

DATABOX

01000100 01100001 01110100 01100001 01100010 01111000

PERSONALISED ADVERTS



Elsie's health insurance is due to expire.

Elsie installs an insurance comparison app on her databox.

The app analyses her home, mobile, location and grocery shopping data.

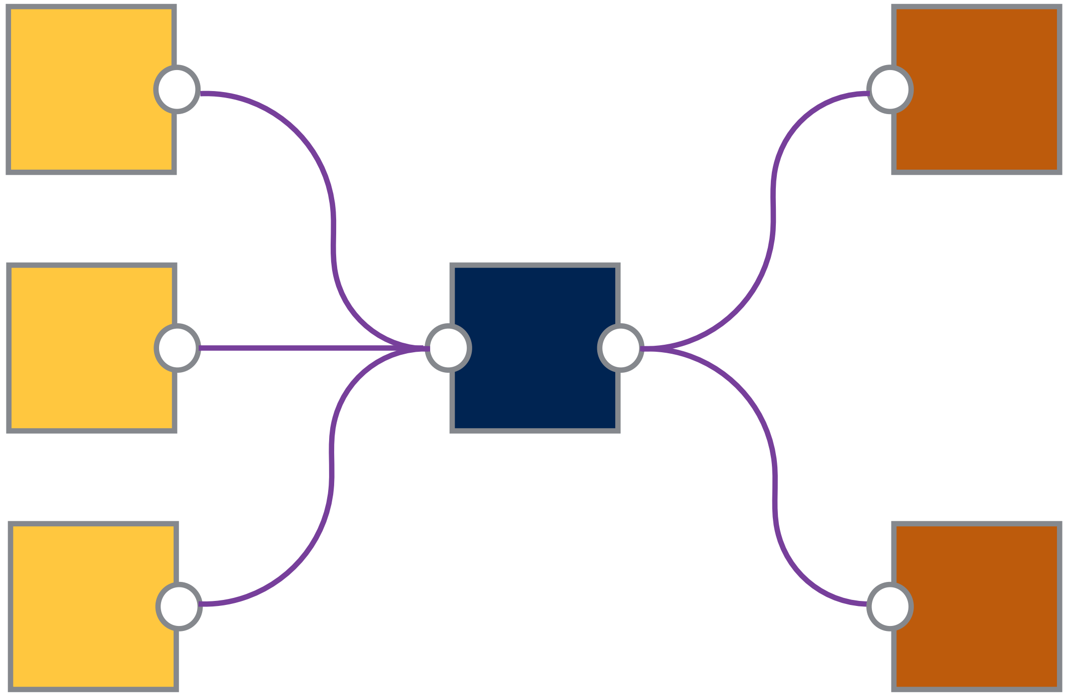
The app discovers that Elsie has an active and healthy lifestyle and offers her a big discount.

DATABOX
01000100 01100001 01110100 01100001 01100010 01111000

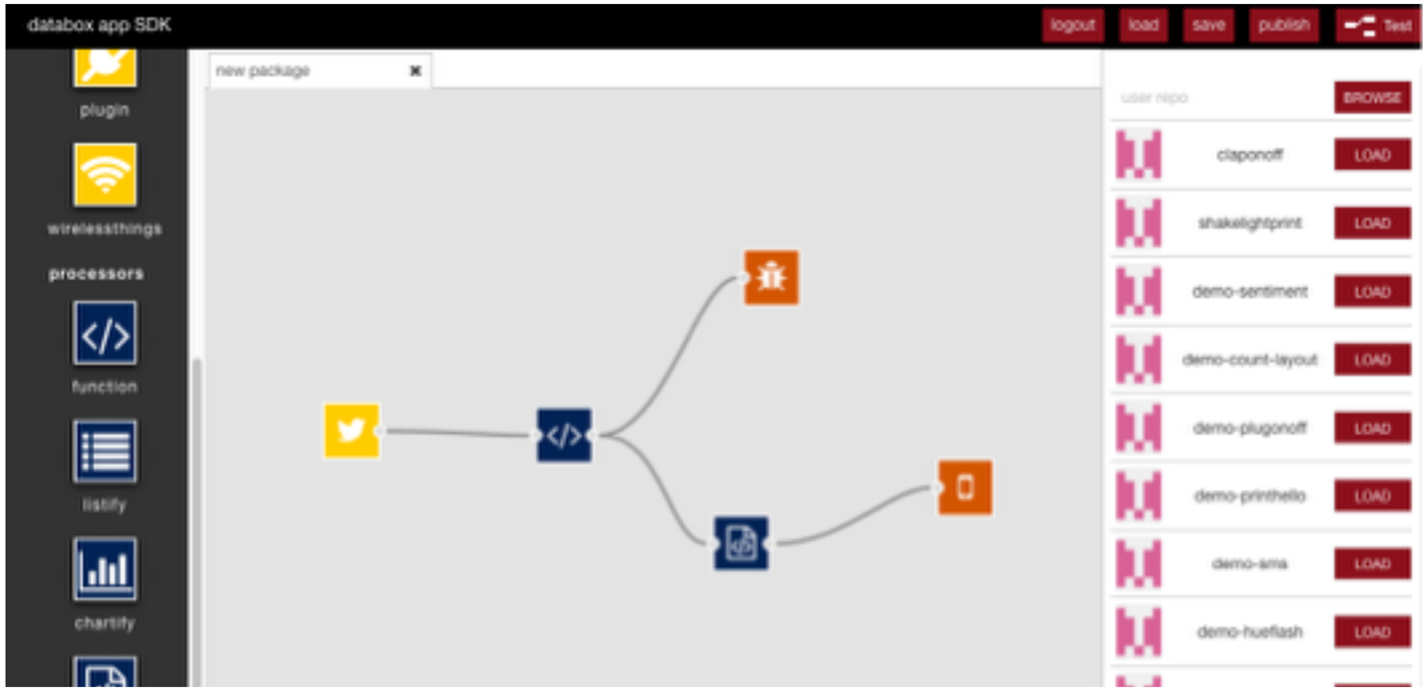
HEALTH INSURANCE

Developing Apps

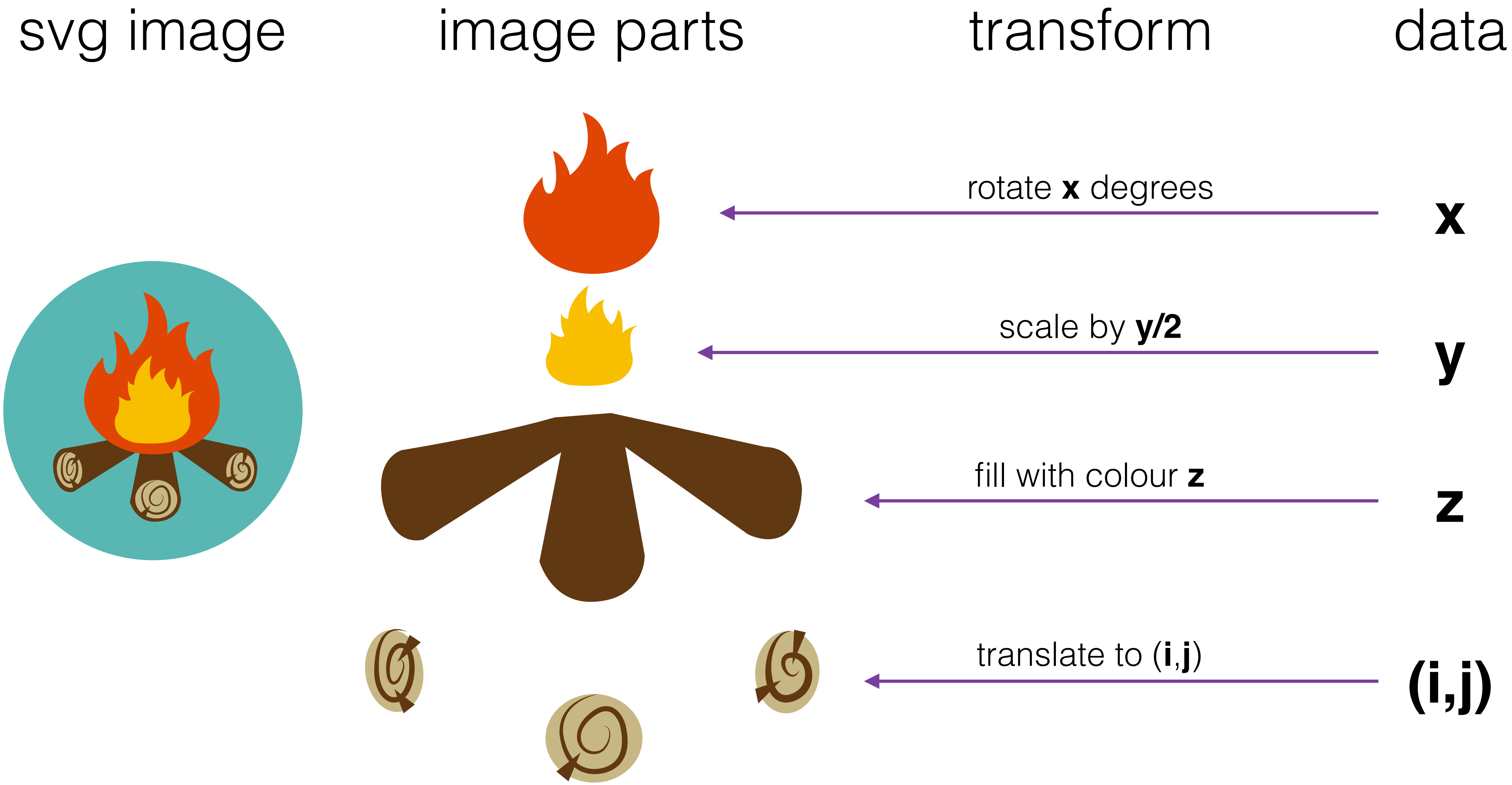
datastores	processors	outputs
hue bulbs	map, reduce	actuate
mobile sensors	filter	display
smart plugs	convert	write to store



- Install and connect existing apps
- Plug together apps and components to customise **your** apps



Rich Visualisations



Physical Interactivity

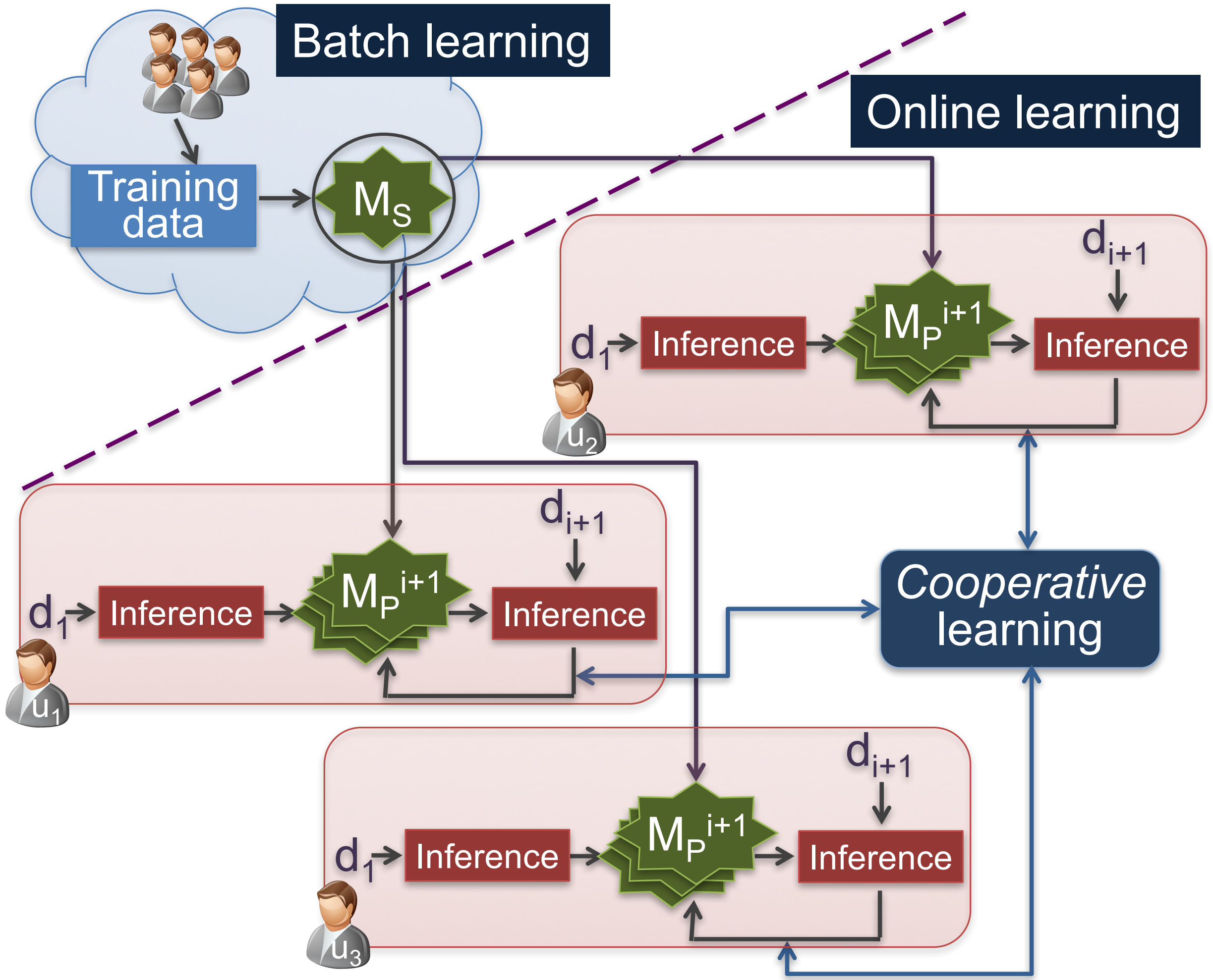
Demo submission to UBICOMP 2017

- Physical devices often easier to reason about
  - Visible; Located; Proximate; Portable
- Physical access control is the norm
  - “The bag of keys” is well understood
- Exploring use of inaudible audio channel to provide physical exchange of virtual capabilities
  - Macaroons, “cookies with caveats”
  - E.g., time limited guest access to actuate lights

Distributed Analytics

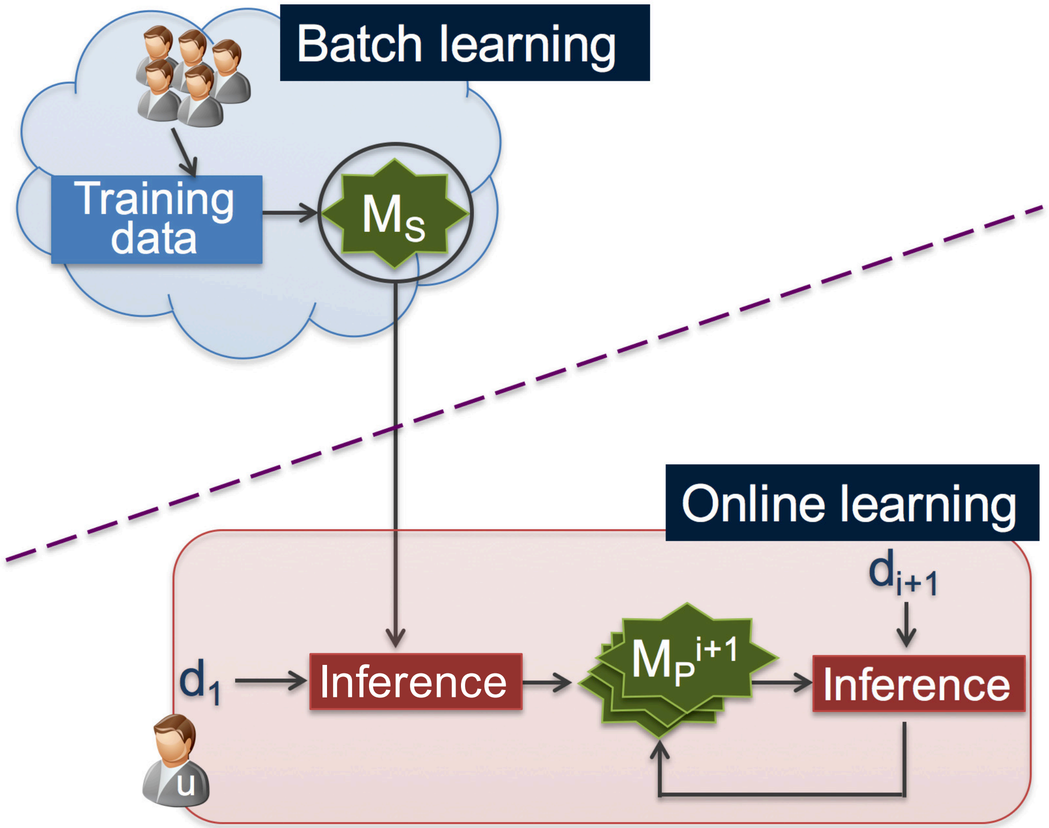
In submission to PoPETS 2017

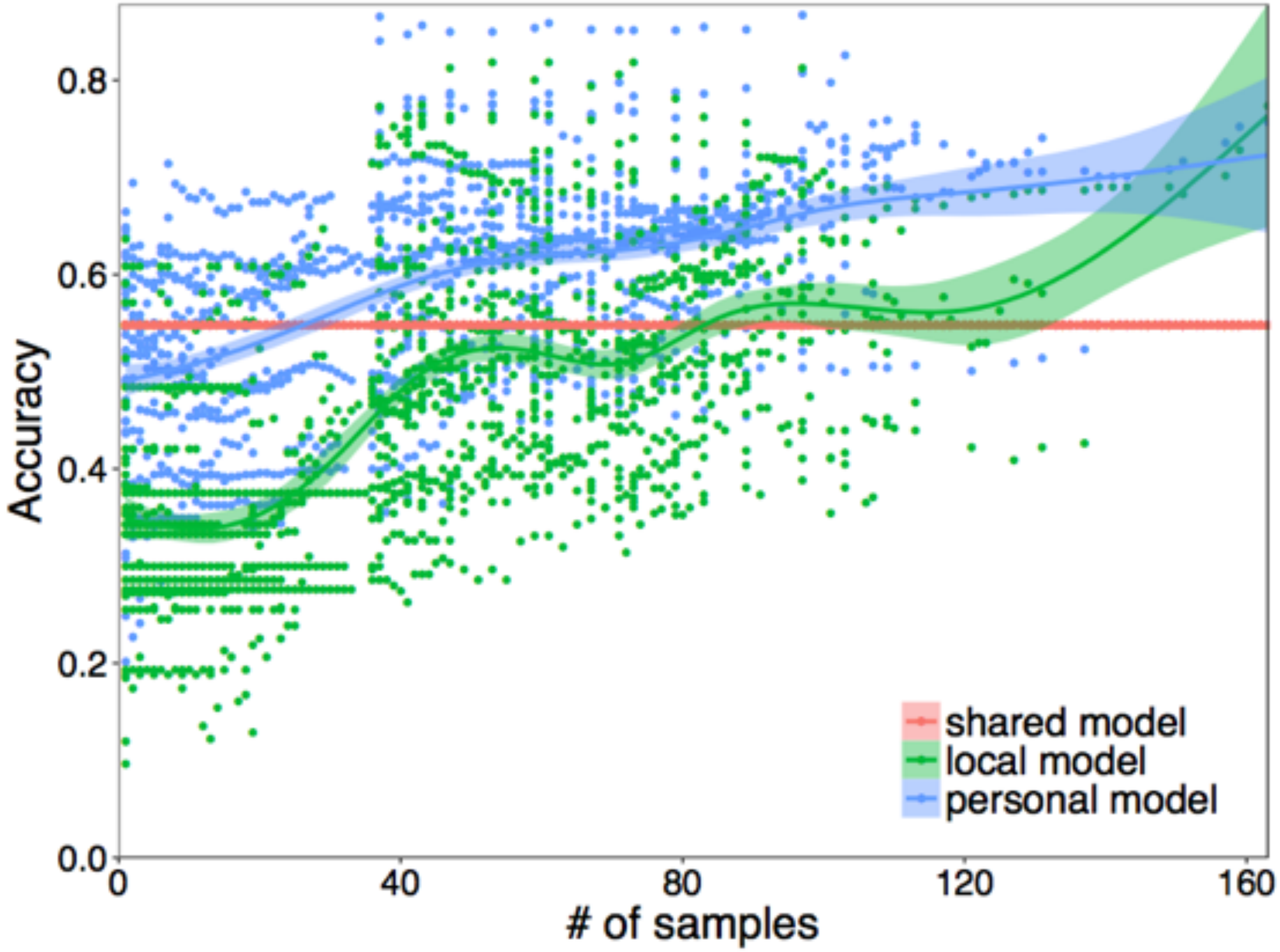
- How to handle scale, heterogeneity, dynamics?
- Subject vs processor driven
  - App stores vs cohort discovery
- Cohort vs individual processing
  - Distributed model building
  - Personal local visualisation



Online Learning

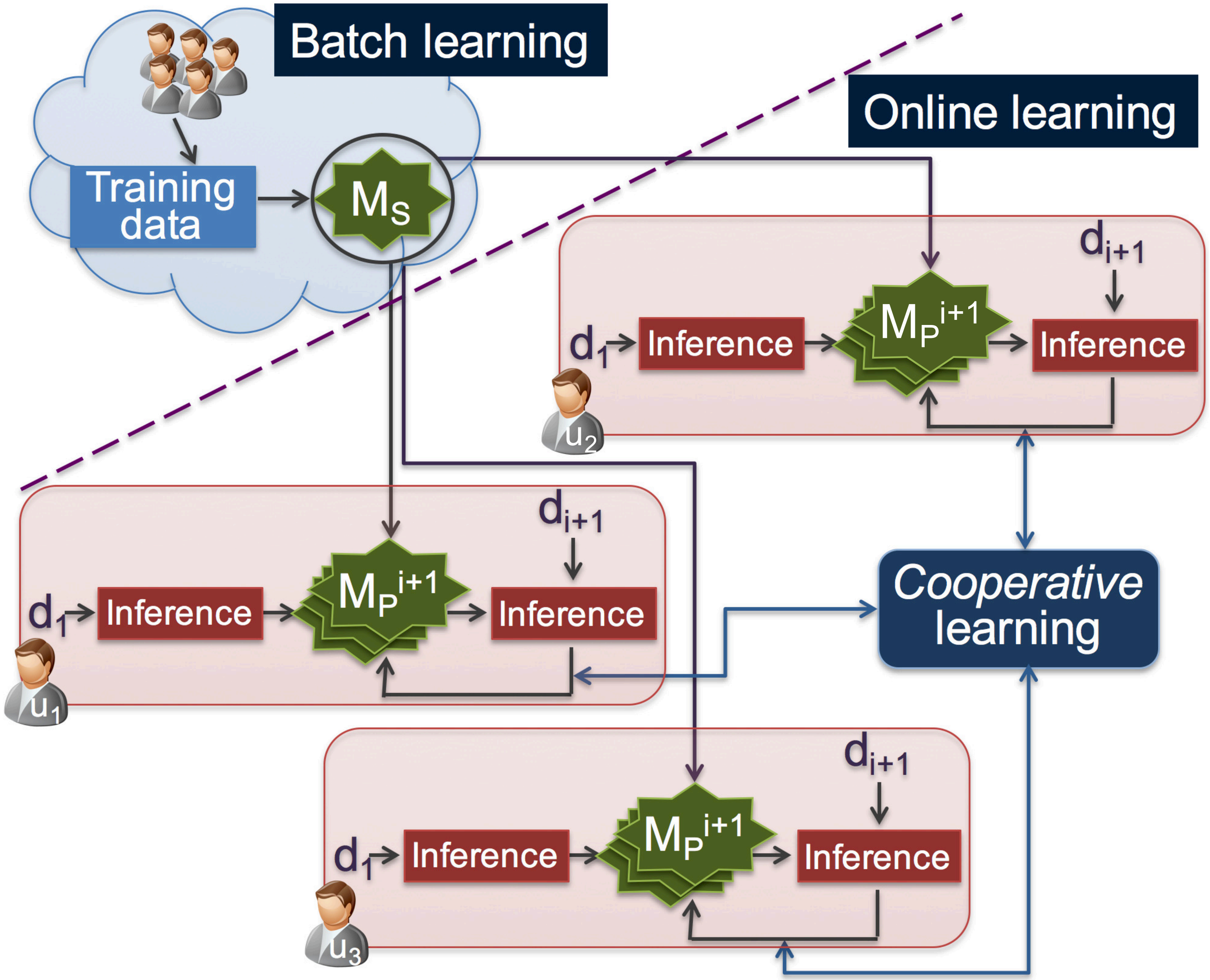
Can we use personal data to improve public, pre-trained ML models?





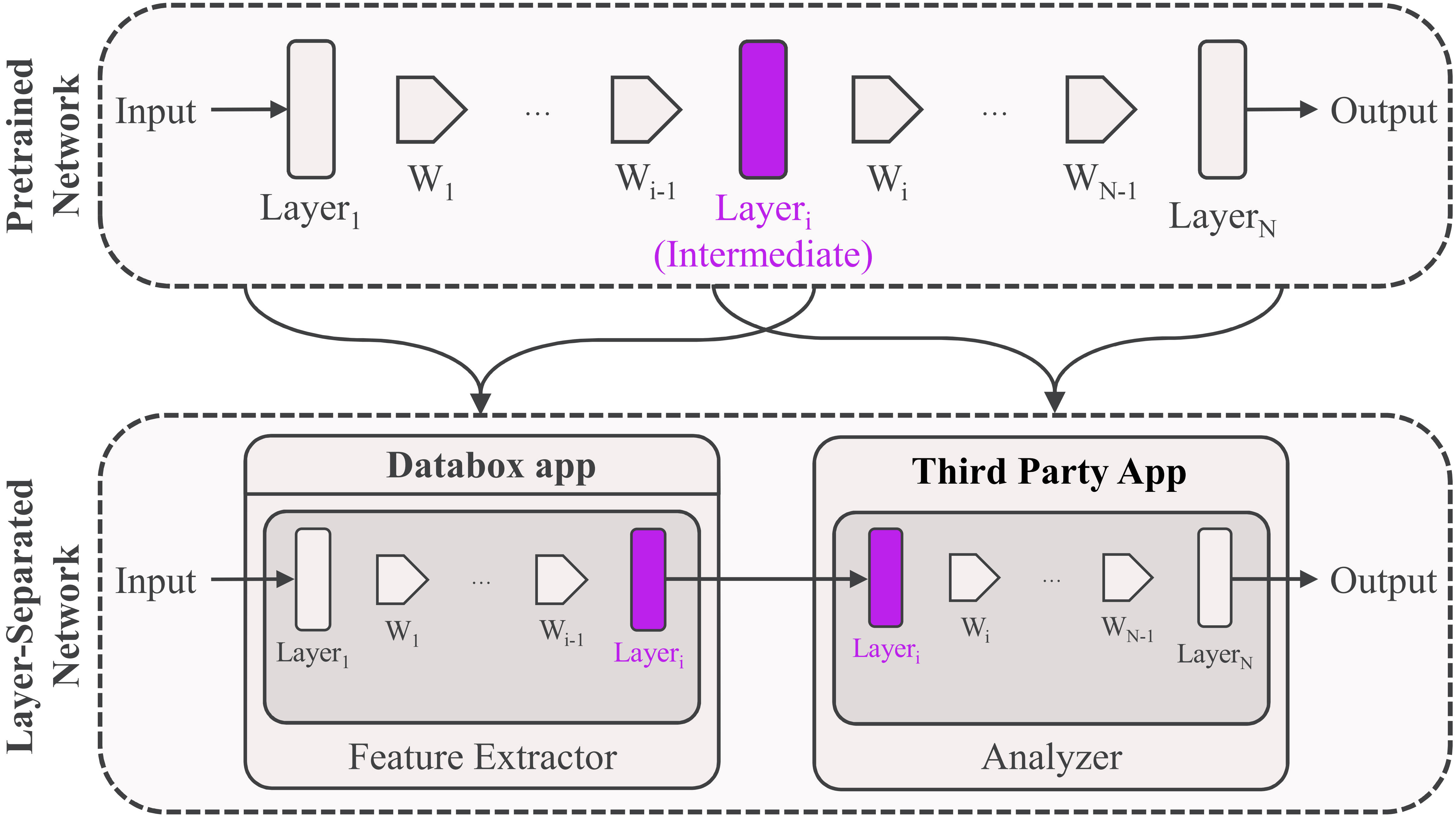
Cooperative Learning

Or train our models cooperatively over distributed users?



Example: Occupancy-as-a-Service

Privacy-Preserving Analytics



Open Source Community Engagement

https://forum.databoxproject.uk/

https://github.com/me-box/

Questions?

http://mort.io/
richard.mortier@cl.cam.ac.uk

https://databoxproject.uk/
https://forum.databoxproject.uk/
http://hdiresearch.org/

McAuley et al, COMSNETS'11
Haddadi et al, Aarhus'15
Crabtree & Mortier, ECSCW'15
Mortier et al, Encyclopaedia of HCI, IDF'16
Mortier et al, CoNEXT CAN'16

Work funded in part by EU FP7 611001, EPSRC EP/N028260/1, EP/N028422/1