

Human-Data Interaction & the Databox

Richard Mortier

@mort__

...with input from

Andy Crabtree, Jon Crowcroft, Hamed Haddadi,
Tristan Henderson, Derek McAuley, Anil Madhavapeddy

Our Digital Footprints

Digital footprints pose **major societal challenges...**

<https://flic.kr/p/ppMdY1>



<https://flic.kr/p/6sdrZB>

...as the same time as an opportunity for **economic growth**

Living in a Big Data World

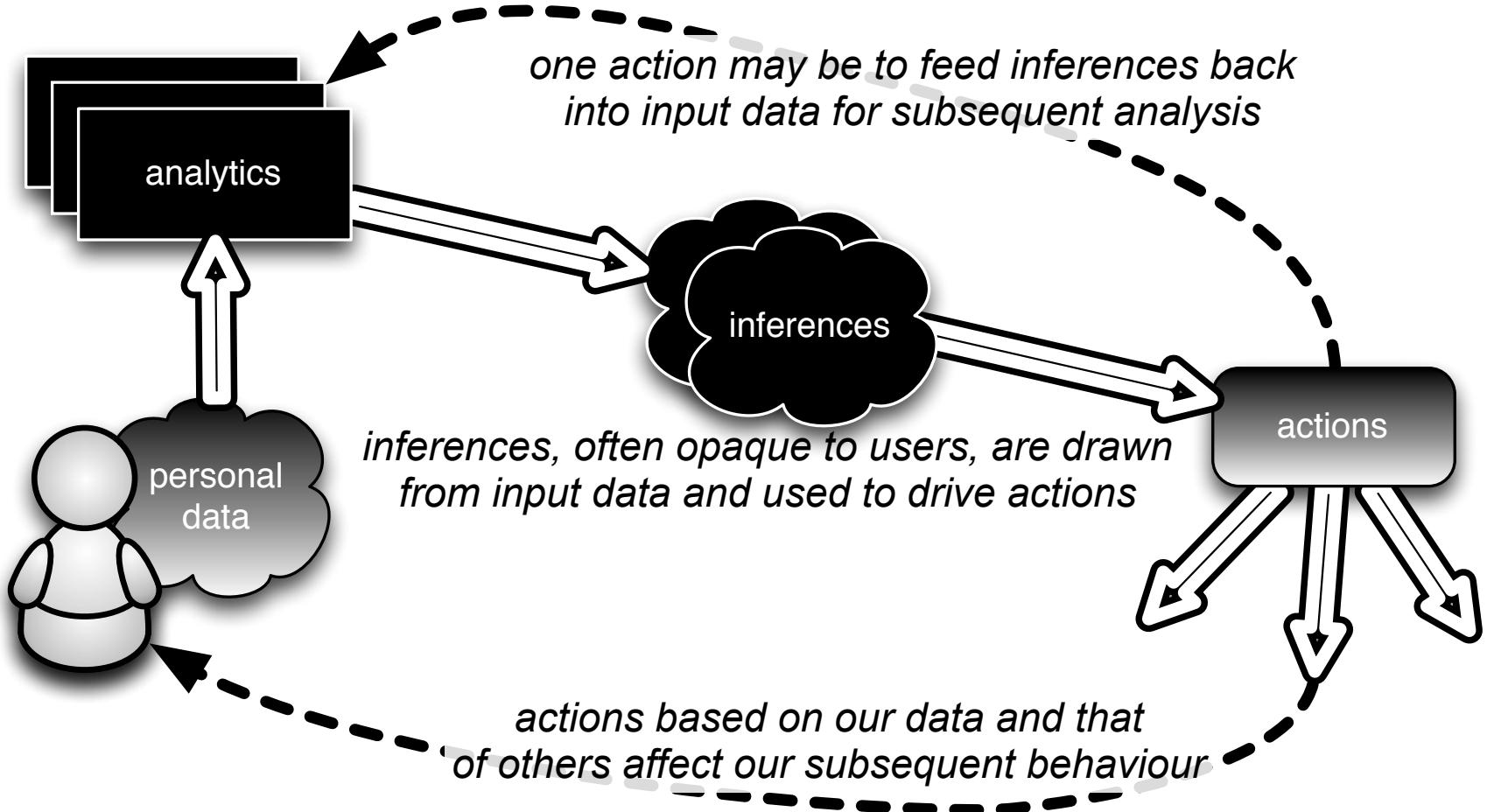
- Intimate information about us is collected and used
- It augments already large, rich data silos
- Never forgetting or forgiving



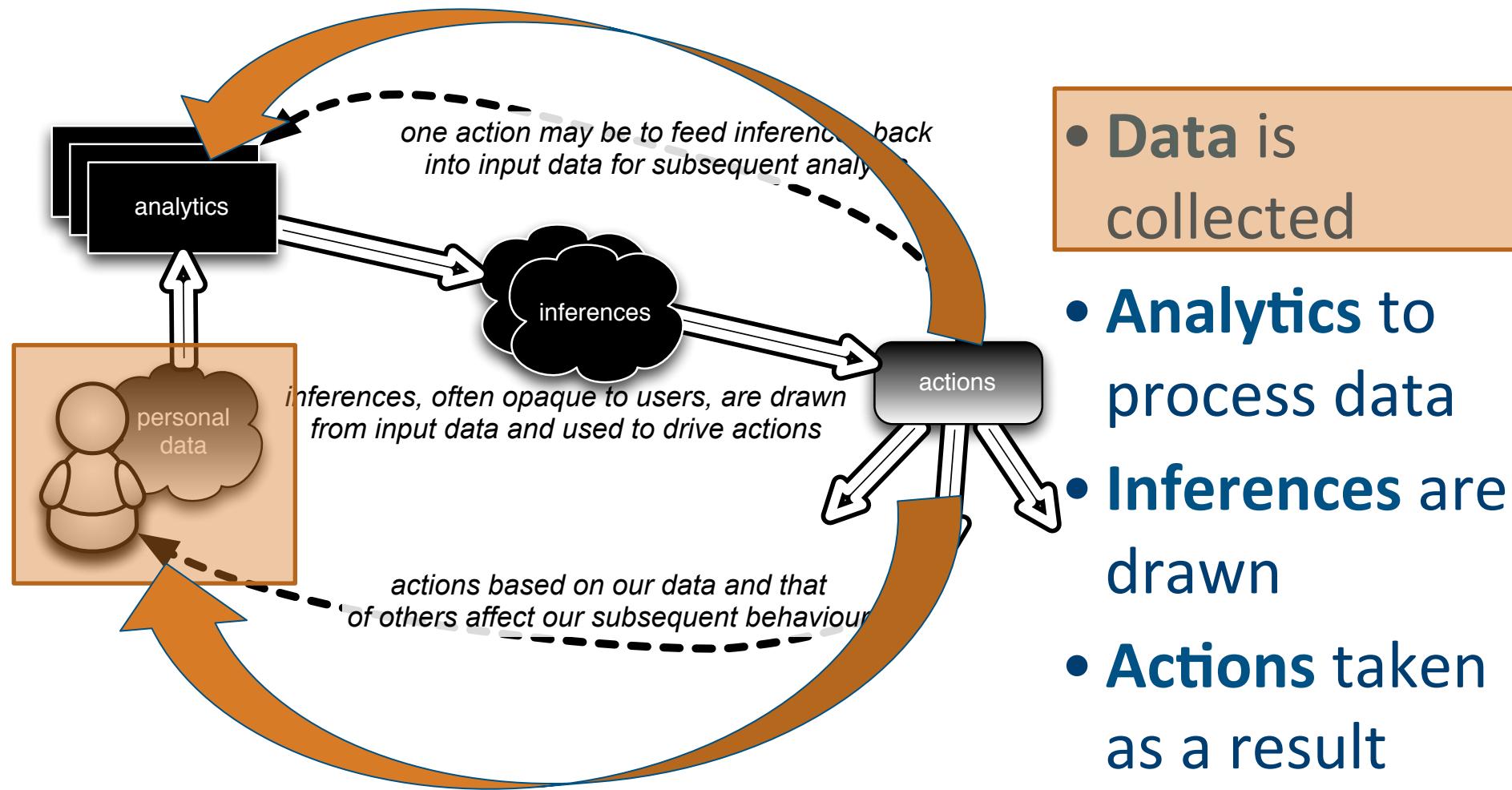
Key Challenge:

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

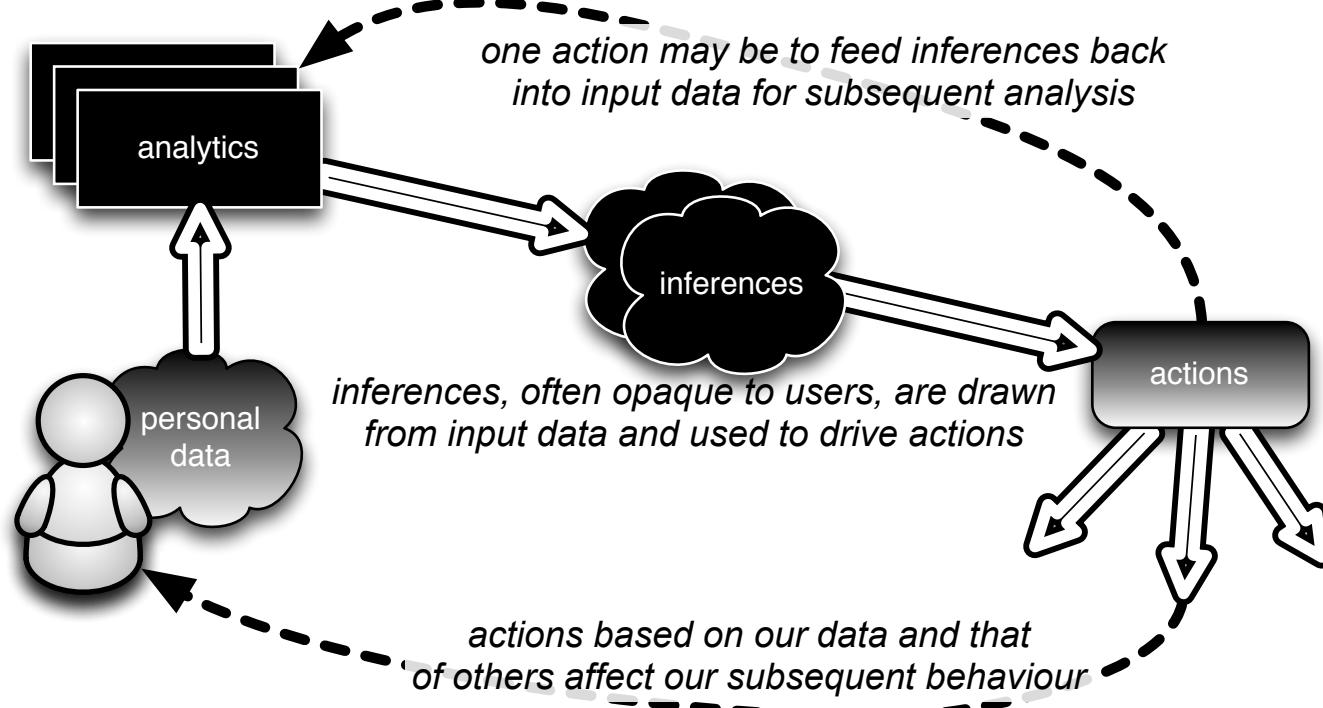
Human-Data Interaction



Human-Data Interaction



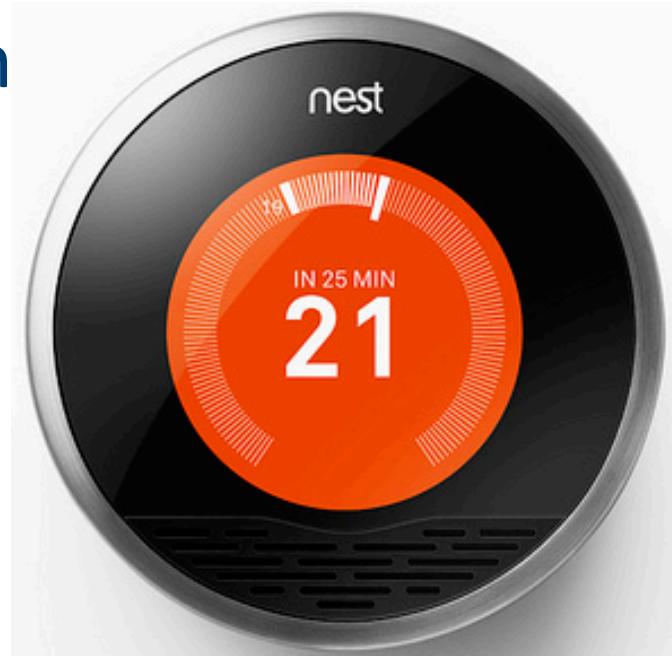
Human-Data Interaction



We believe current systems lack
Legibility, Agency, Negotiability

Legibility

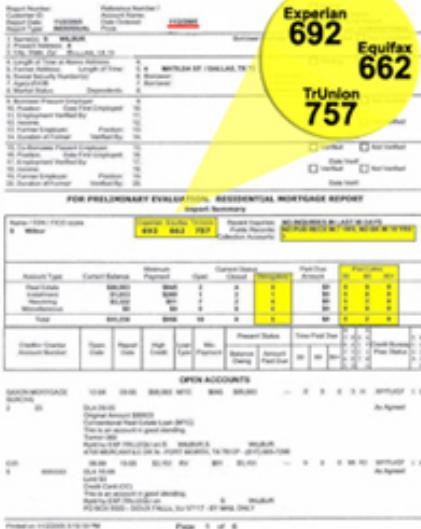
- Visualisation & comprehension
- E.g., Nest thermostat
 - Simple information display
 - Supports many interaction modalities
 - Hides details of internal processes



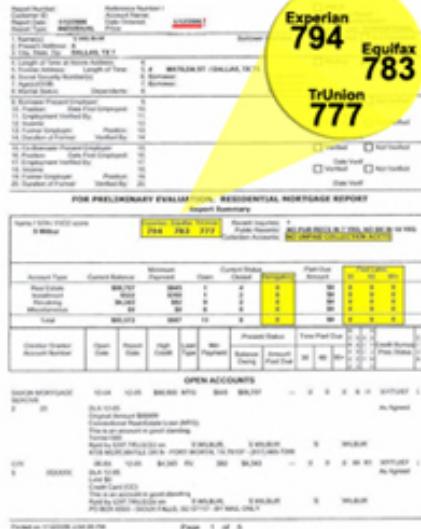
<https://flic.kr/p/azwi7q>

Lack of Legibility

Credit Report - BEFORE



Credit Report - AFTER



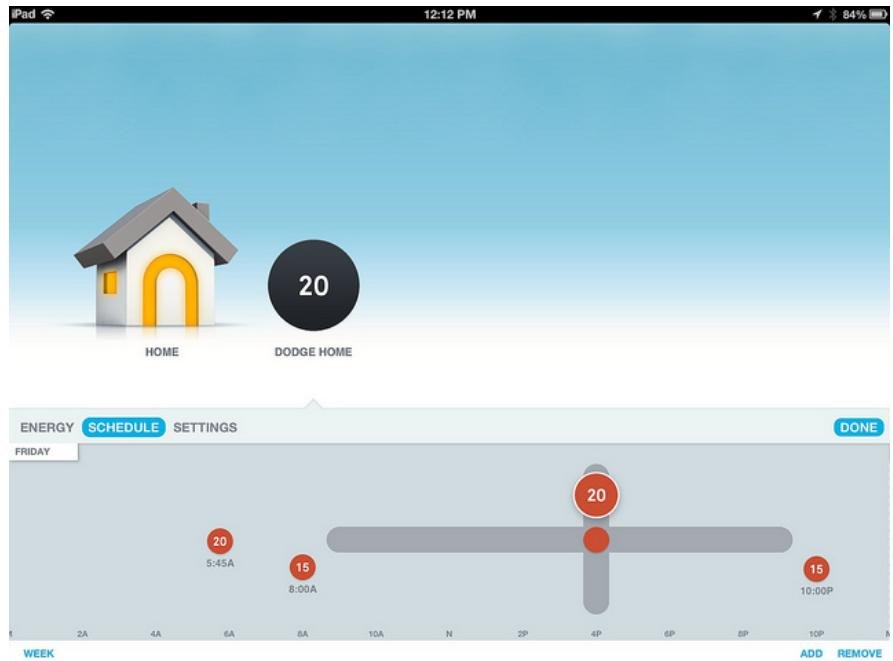
- We are unaware of
 - the many **sources of data** collected about us,
 - the **analyses performed** on this data, and
 - the **implications** of these analyses.

<https://flic.kr/p/6thmfN>

- E.g., Computation of credit scores

Agency

- Capacity to act
- E.g., Nest Thermostat
 - Learns a schedule, but
 - Supports user override, by
 - Setting desired temperature on-demand



<https://flic.kr/p/e3oH3k>

Lack of Agency



Tricia Wang

annoyed at academic Facebook researchers, SET YOUR PRIVACY SETTINGS to FRIENDS if you don't want people trolling around with y info: <http://tinyurl.com/2r8j9>.
Updated just a moment ago [edit](#)

Networks: UCSD Alum '07
Sex: Male
Interested In: Women
Men
Looking For: Whatever I can get
Hometown: Tricia Love, Akrotiri
Political Views: Very Conservative
Religious Views: Fundamentalist Member of the Congregation Tricia!

<https://flic.kr/p/4eQXTn>

- We are unaware of
 - the means we have to affect data collection,
 - the means we have to affect data analysis,
 - if they even exist, and we know enough to want to employ them
- E.g., Use of purchase details to profile your propensity to risk and sell this to an insurance agency

Negotiability

- Support for the dynamics of interaction
- E.g., Nest Thermostat
 - Provides means to inspect and edit the schedule it has learnt
 - Continually updates learnt behavior to adapt to changes in context
 - Based on context-dependent patterns of past user interaction



<https://flic.kr/p/i8cHvi>

Lack of Negotiability

Even given

- we know the data collected and analyzed about us, and
- we understand how to enact choices over these

We're still trapped by current systems and services

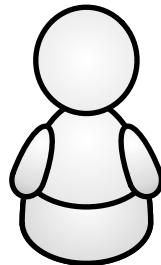
- Binary accept/reject of terms
- Cannot subsequently modify or refine our decisions
- Cannot easily correct data or inferences held about us

The screenshot shows the 'Ads Settings' page at www.google.com/ads/preferences. The page is titled 'Settings for Google ads' and explains that ads enable free web services and content. It shows two main sections: 'Ads on Google' and 'Google ads across the web'. Under 'Ads on Google', there are four categories: Search (blue magnifying glass icon), Gmail (red envelope icon), YouTube (red play button icon), and Maps (green location pin icon). Under 'Google ads across the web', there is one category: 'Google ads across the web' (blue square icon with a cursor). Below these sections, there are four rows of user information:

Setting	Value	Source
Gender	Male	Visit your Google Profile
Age	35-44	Visit your Google Profile
Languages	N/A	
Interests	Bollywood & South Asian Film, and 14 more	Edit From your previous activity on Google
	Action & Adventure Films, and 24 more	Edit Based on the websites you've visited

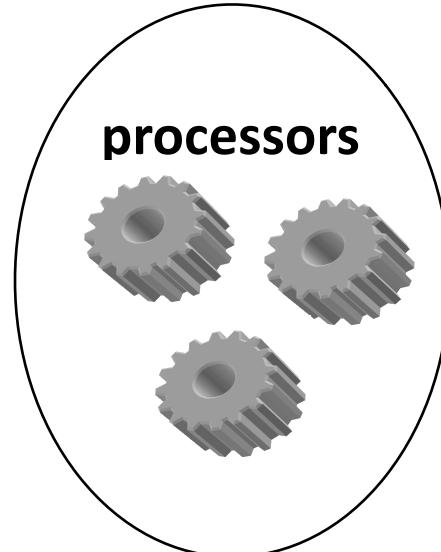
Dataware: The Actors

subject

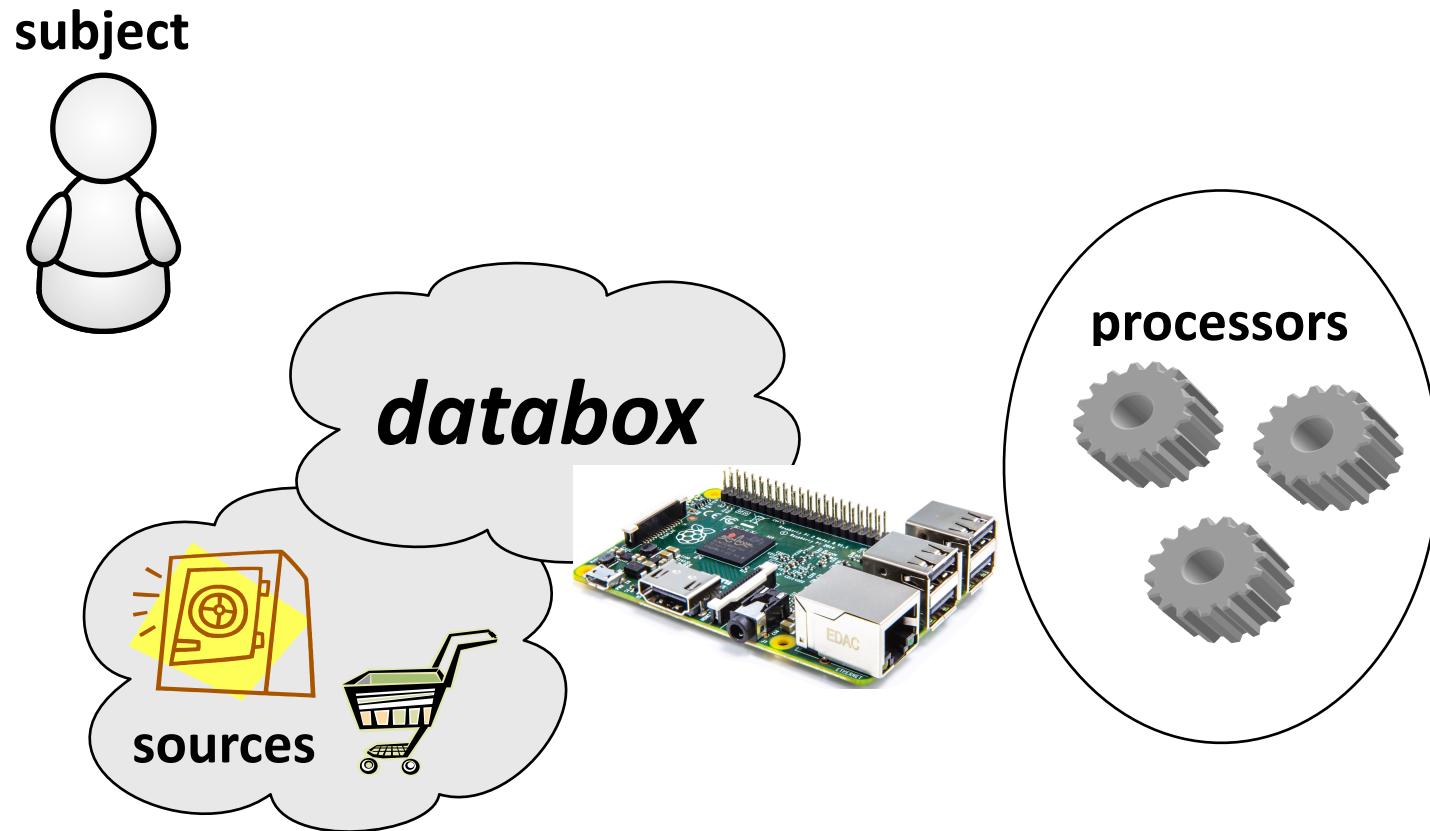


sources

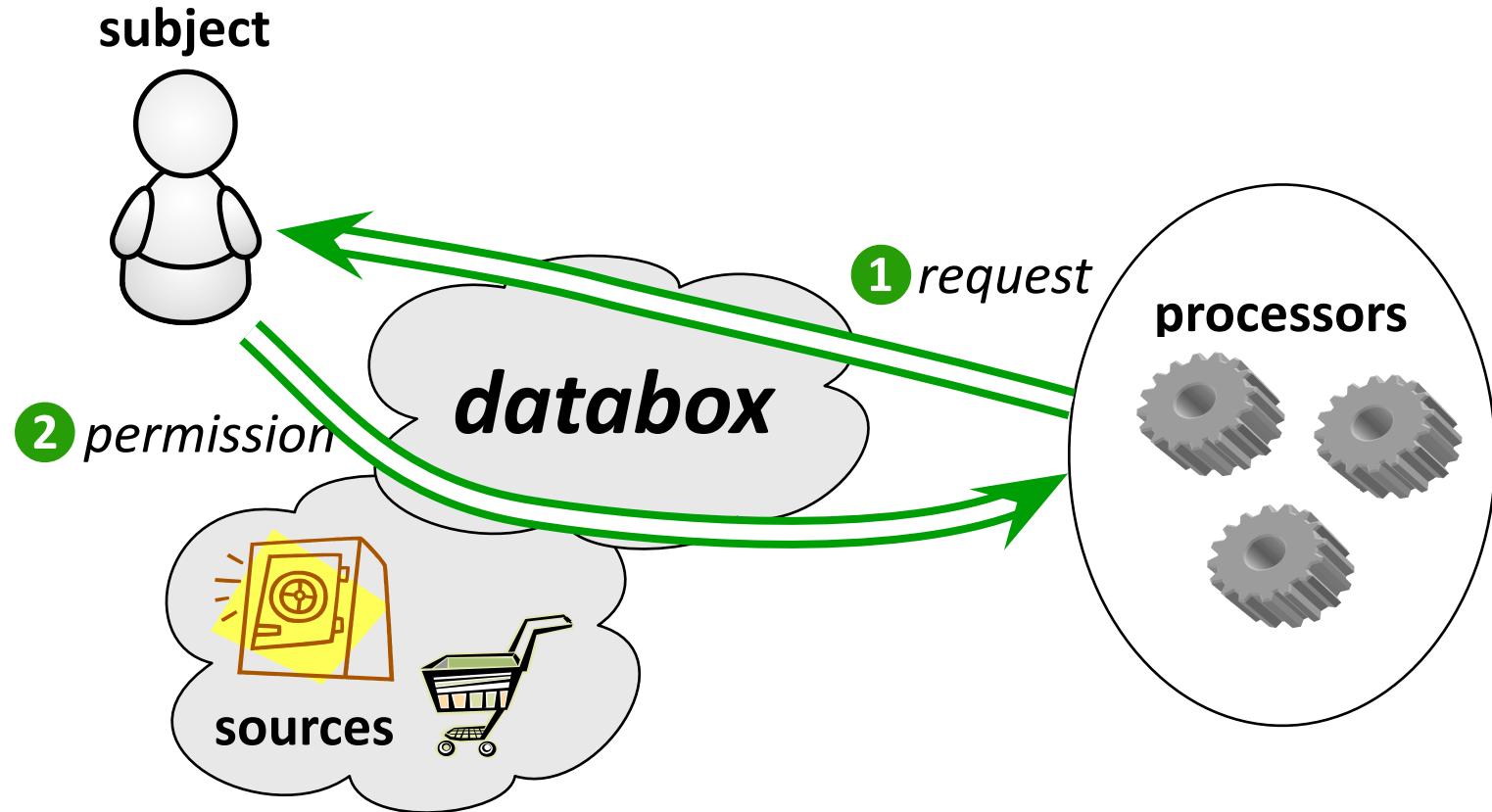
processors



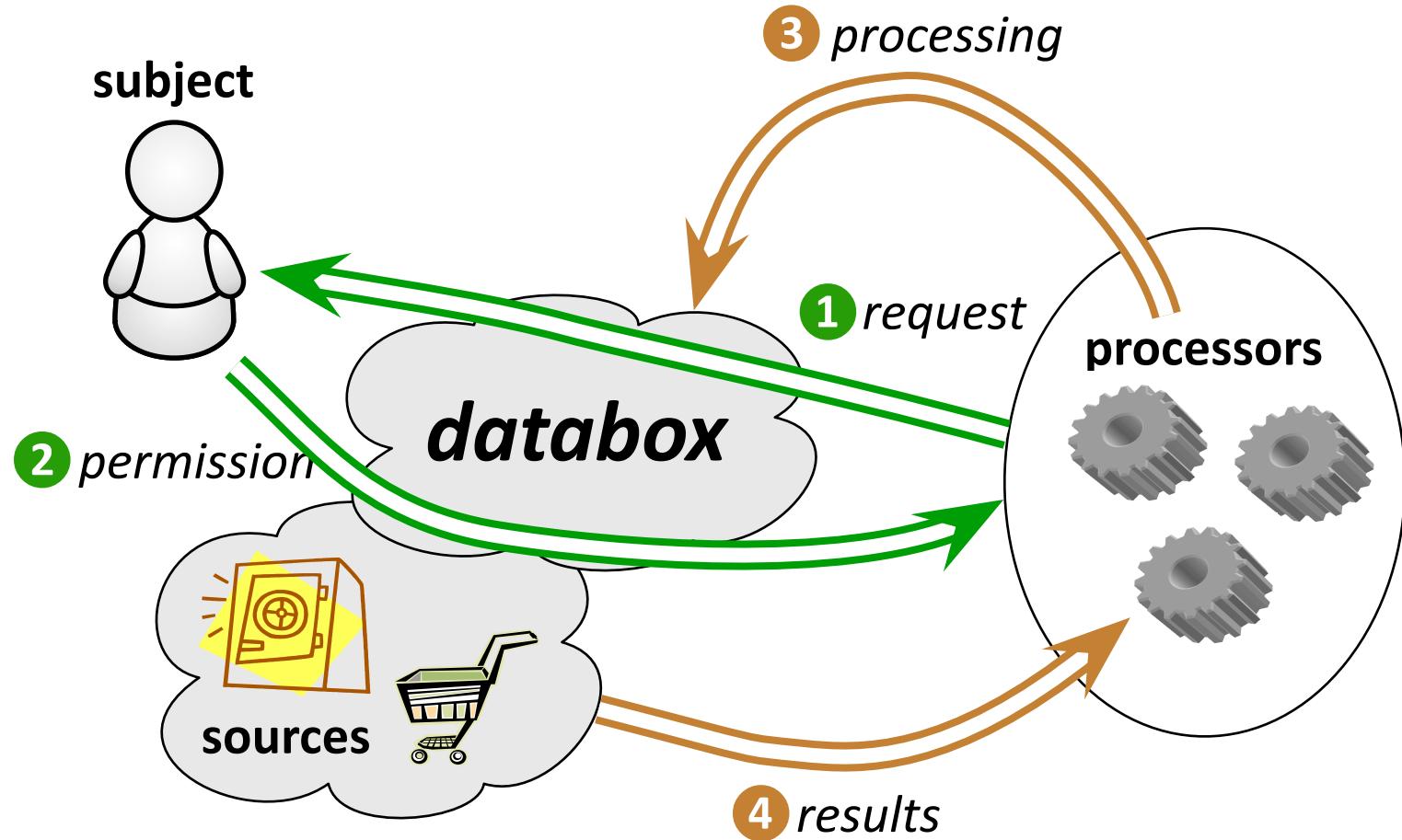
Dataware: Implementing HDI



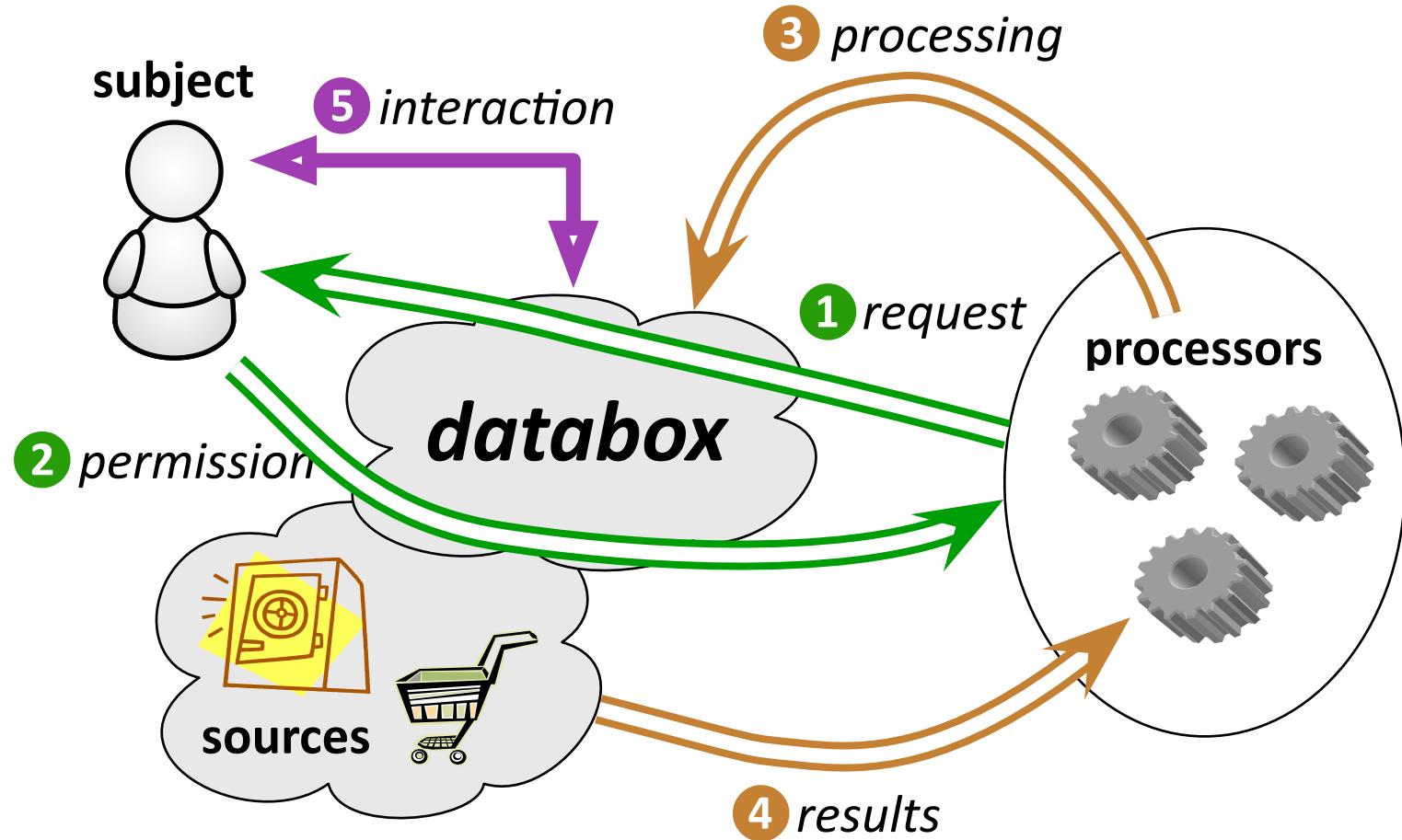
Dataware: Legibility



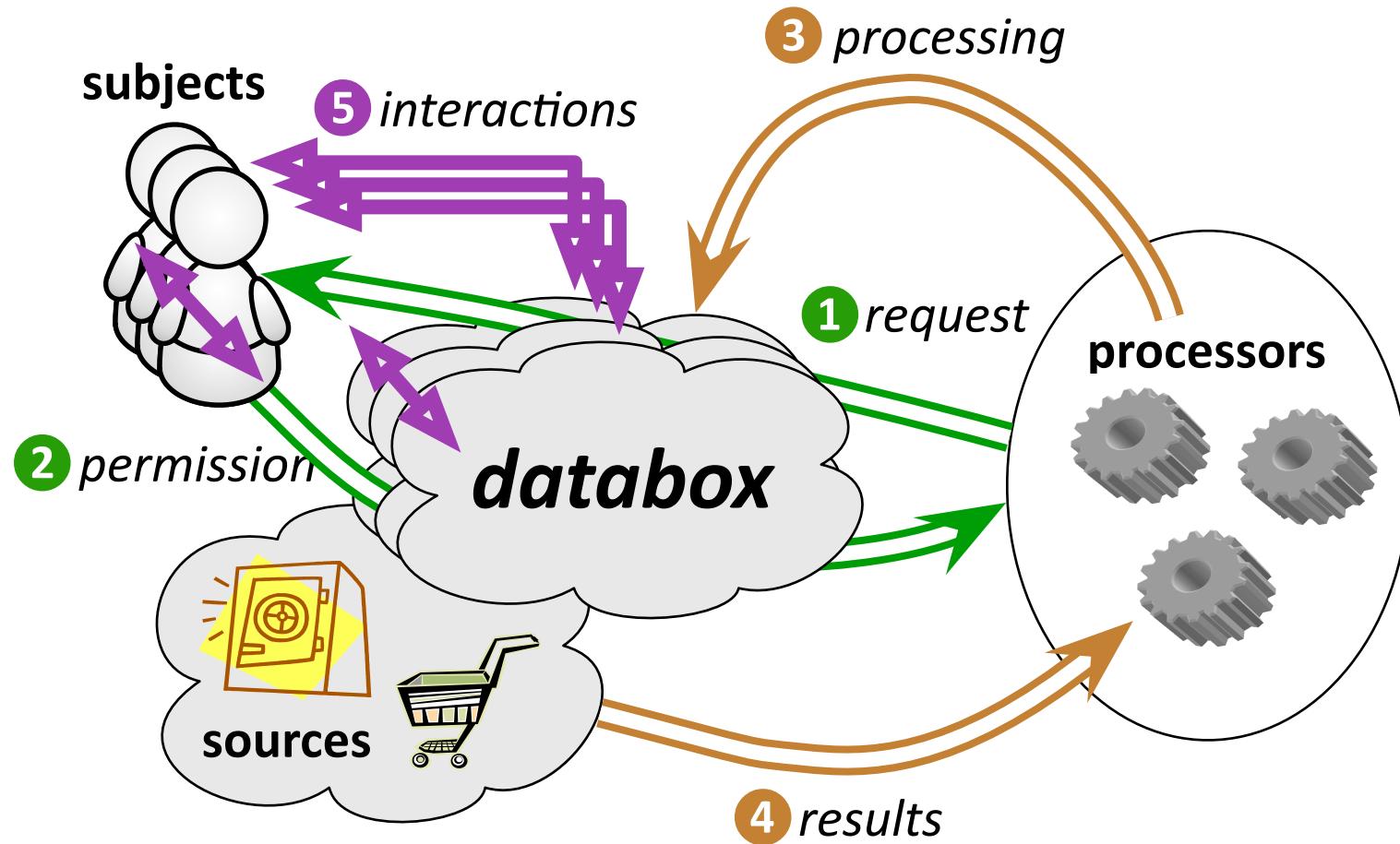
Dataware: Agency



Dataware: Negotiability



Dataware: Constructing Interaction



Dataware: Constructing Interaction

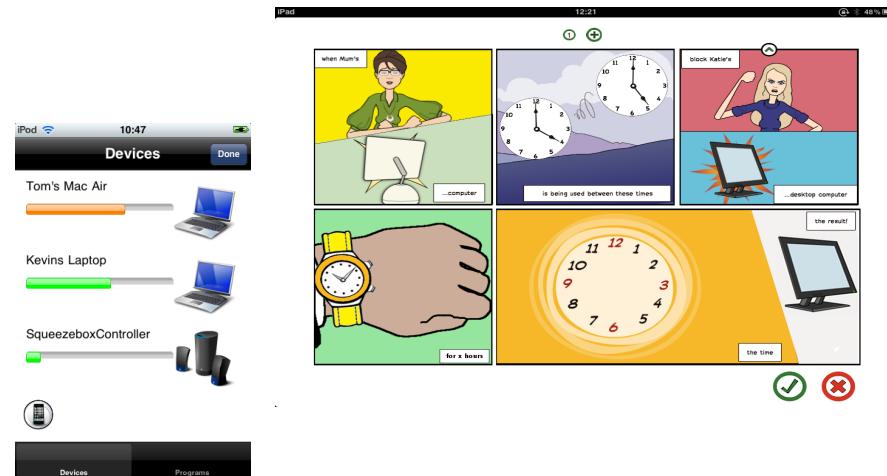
- Numerous proposed interaction models
 - E.g., pay-per-use
- Little about how to actually provide for it
- Dataware one such proposal
 - Accountable transaction between parties in terms of request, permission, audit
- But there's a lot more to consider here...

Data as a Boundary Object

- Contextual nature – plastic adaptation to need
- E.g., Credit card receipt
 - Consumer's proof of **payment**
 - Bank's proof of a **valid transaction**
 - Supermarket's proof that **the bank should pay them**
- Inherently relational and thus social
 - Not so much 'me' or 'you' as 'us'
 - Very little is so private that it involves no-one else

Not Even as Simple as a Home Network

- Disambiguation can't be delegated to a nominated householder/cohort
 - Too many relational issues wrapped up in this
 - Old, young; Parents, children; Colleagues, friends, lovers
- Not even just about my **vs** our data
 - We may not agree



HDI: So Where's the Interaction?

- Request and processing occur as if in a black-box
 - Can't tell where it's got to, what's going on
 - Status within the arrangement
- Requests, permissions and audit logs
 - Mechanisms of coordination within the field of work
 - Order but do not articulate the field of work
- Real world data sharing is **recipient designed**
 - Shaped by people with respect to the relationship they have with the parties implicated in the act of sharing

Articulation Work

- Dataware subject is engaged in cooperative work
 - Interdependence between subject, processor, perhaps other subjects
- Thus these activities must be meshed together
 - E.g., Schmidt (1994)
 - *maintaining reciprocal awareness* of salient activities within a cooperative ensemble
 - *directing attention* towards current state of cooperative activities
 - *assigning tasks* to members of the ensemble
 - *handing over* aspects of the work for others to pick up

Interactional Challenges for HDI

User Driven Discovery

- What is discovered? By whom? Under whose control?
- Need for metadata usage analytics
- Empowering subjects: app stores
- Permissions, social ratings and exchange

From *My Data* to *Our Data*

- Delegating and revoking control
- Editing, viewing, sharing
- Group management, negotiated collection and control

Legibility of Data Sources

- Visualisation of own data, impact of others' data
- Present and future public data
- What you have, what others want
- Recipient Design: editing of data; control of presentation to processors

Salient Dimensions of Collaboration

- Transitivity: to whom is data passed, for what purpose
- Tracking and treatment

Thematic Areas for HDI

- **Personal data discovery**, including meta-data publication, consumer analytics, discoverability policies, identity mechanisms, and app store models supporting discovery of data processors
- **Personal data ownership and control**, including group management of data sources, negotiation, delegation and transparency/awareness mechanisms, and rights management
- **Personal data legibility**, including visualisation of what processors would take from data sources and visualisations that help users make sense of data usage, and recipient design to support data editing and data presentation
- **Personal data tracking**, including real time articulation of data sharing processes (e.g., current status reports and aggregated outputs), and data tracking (e.g., subsequent consumer processing or data transfer)

Questions!

<http://mort.io/>

richard.mortier@cl.cam.ac.uk

<http://hdiresearch.org/>

[McAuley, Mortier & Goulding, COMSNETS'11](#)

[Haddadi et al, arXiv'14](#)

[Mortier et al, SSRN'14](#)

[Crabtree & Mortier, ECSCW'15](#)



User-Centred Infrastructure



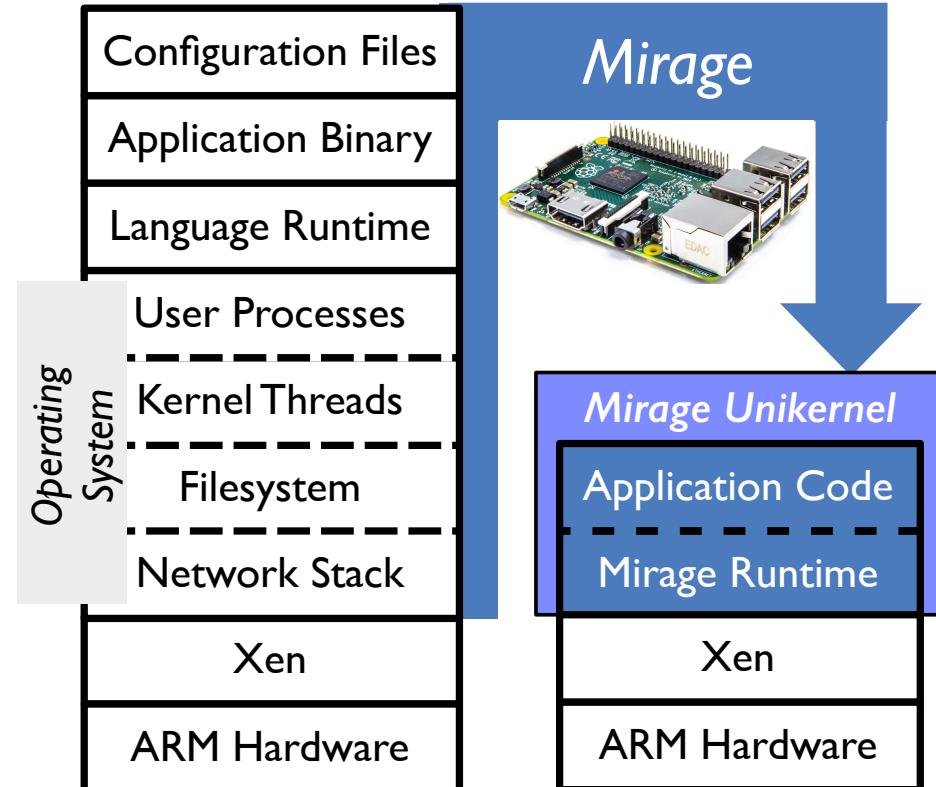
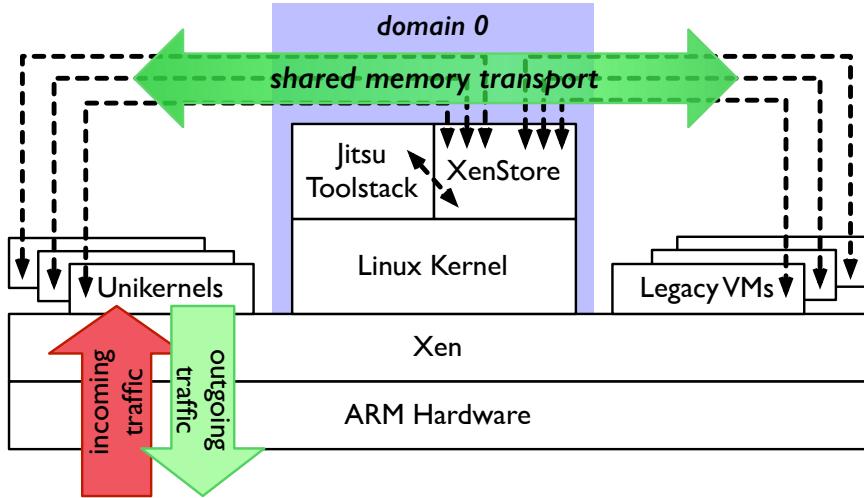
Stable, hidden, shared vs
Dynamic, exposed, intimate



Personal Clouds

<http://openmirage.org/>

- We should operate our own infrastructure
 - ...not abrogate our lives to “the cloud”
- Redesign OS infrastructure for **network services** to be run by **non-expert admins**



Signposts

<http://nymote.org/>

- How can we enable our devices to interconnect?
 - ...directly, without giving everything up to the cloud!



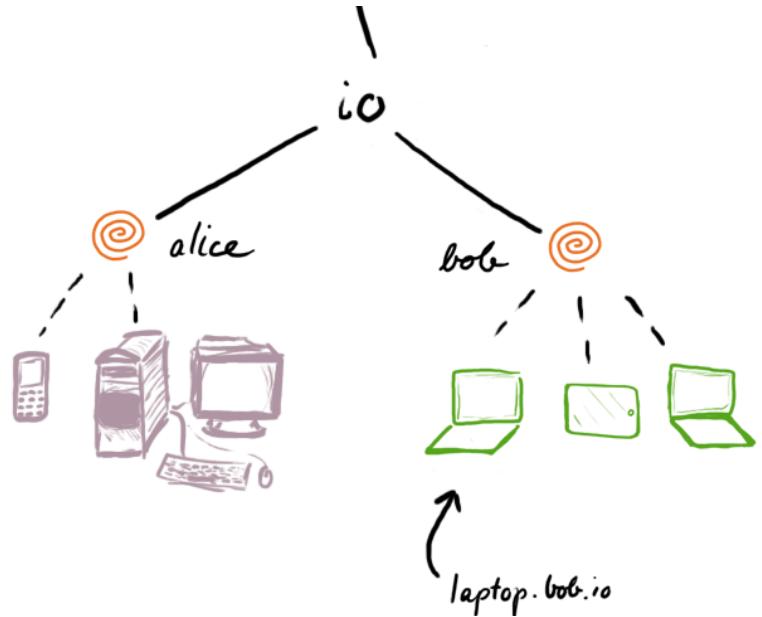
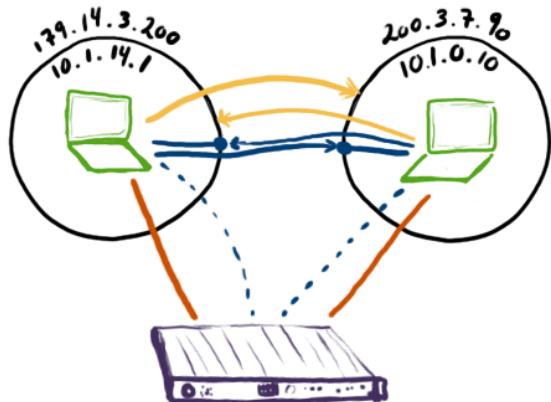
Direct connection



Nat Punching



OpenVPN



- Strong identity
- Orchestration
- Security and protection