

# Human-Data Interaction & the Databox

Richard Mortier

@mort\_\_

...with input from

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

horizon  
DIGITAL ECONOMY RESEARCH

Networks & Operating Systems  
SRG, Computer Laboratory

# 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

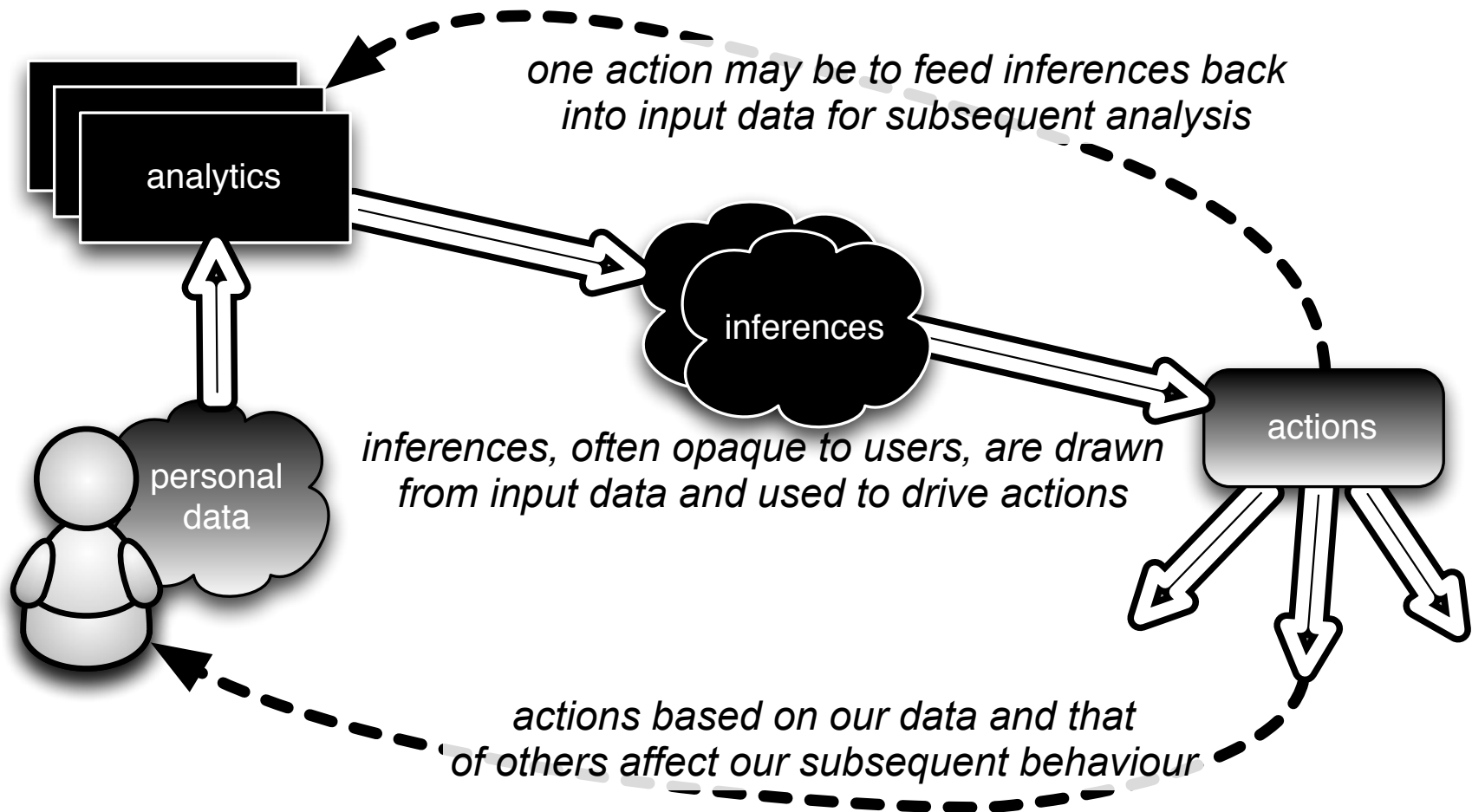
- Intimate information about us is collected and used
- It augments already large, rich data silos
- Never forgetting or forgiving



## Key Challenge:

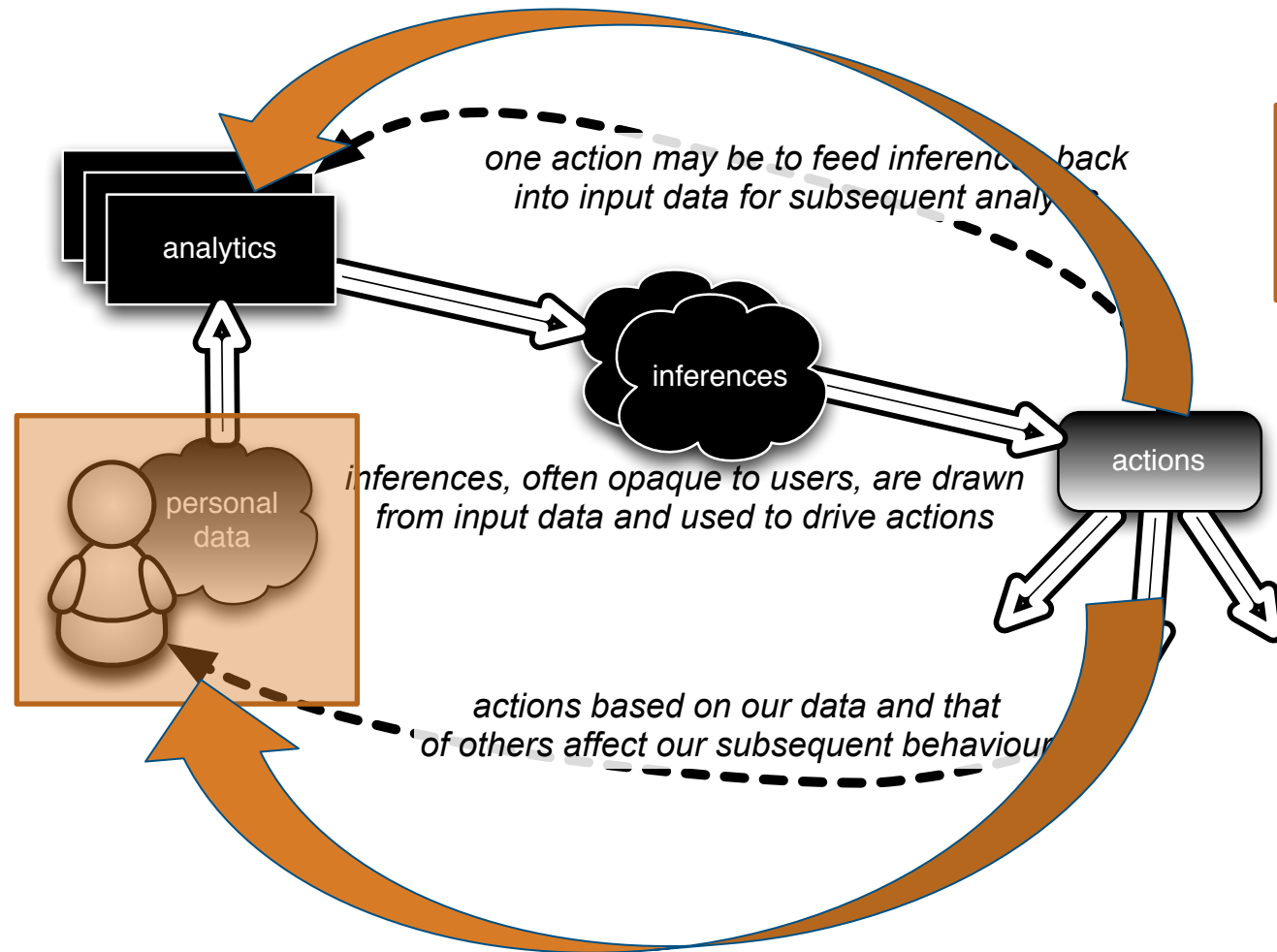
How do we enable individuals to control collection and exploitation of both *their data* and *data about them*?

# Human-Data Interaction



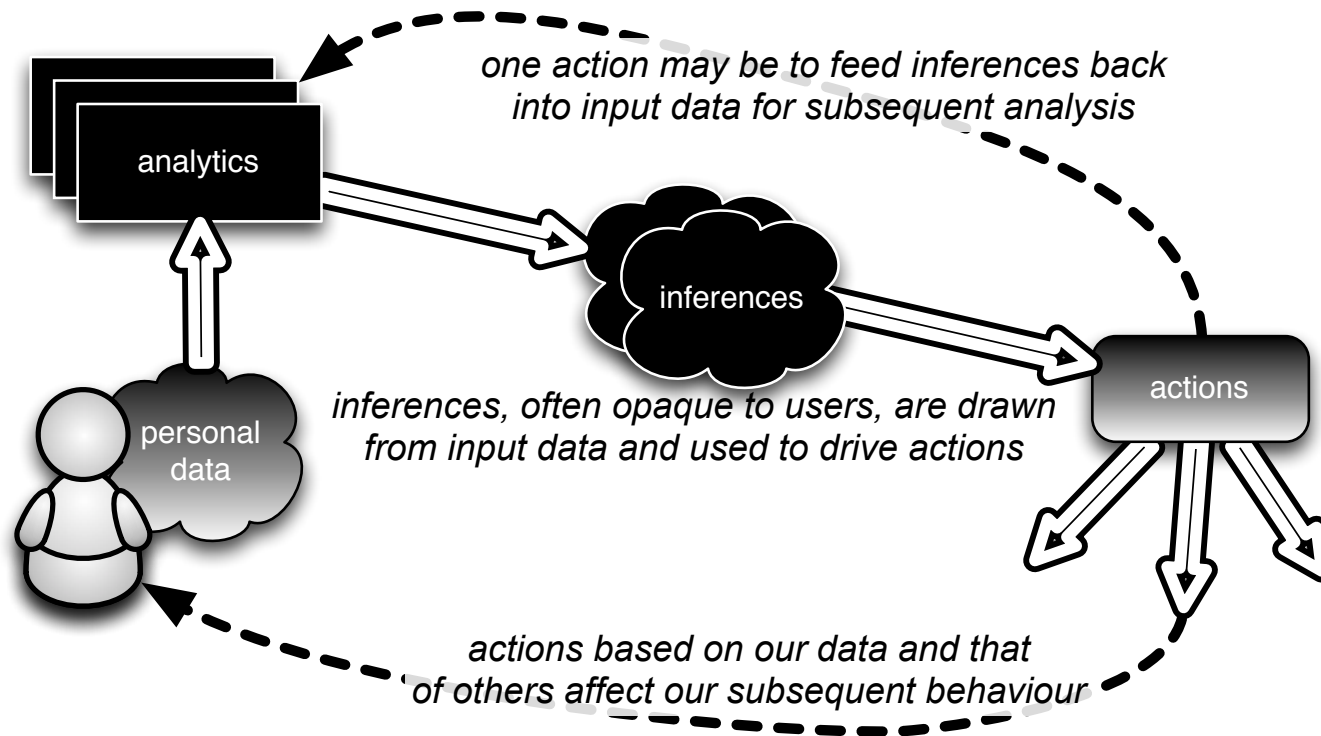


# Human-Data Interaction



- Data is collected
- Analytics to process data
- Inferences are drawn
- Actions taken as a result

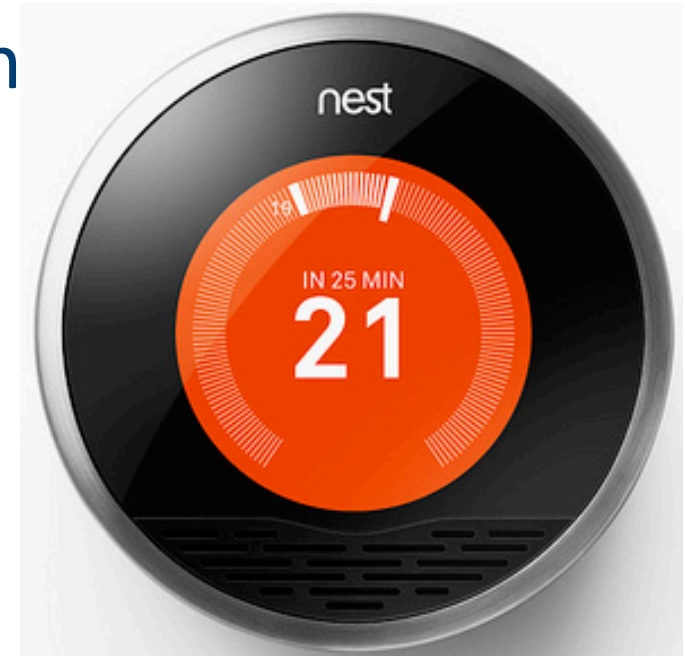
# Human-Data Interaction



We believe current systems lack  
**Legibility, Agency, Negotiability**

# Legibility

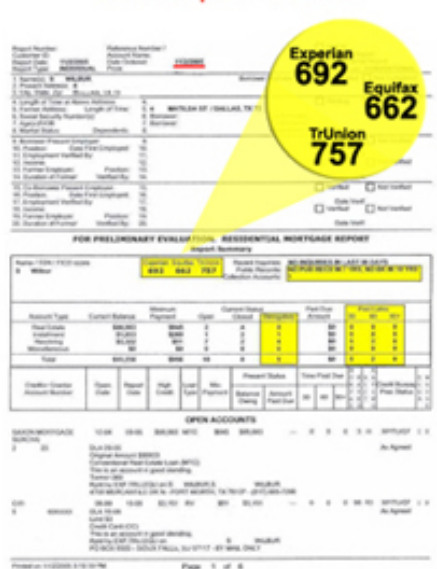
- Visualisation & comprehension
- E.g., Nest thermostat
  - Simple information display
  - Supports many interaction modalities
  - Hides details of internal processes



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

# Lack of Legibility

Credit Report - BEFORE



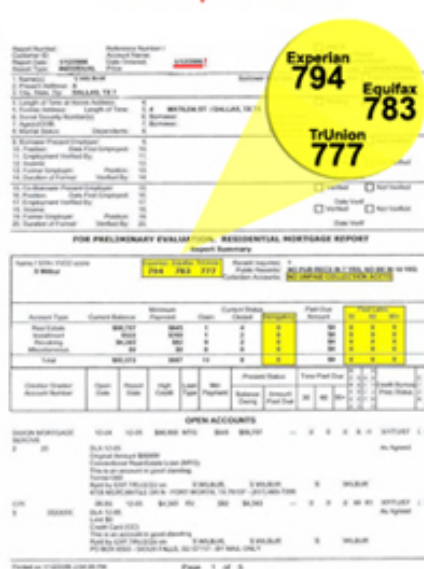
A credit report from Experian, Equifax, and TransUnion. A yellow circle highlights the credit scores: Experian 692, Equifax 662, and TransUnion 757. The report includes sections for account information, payment history, and open accounts.

Account Type	Current Balance	Minimum Payment	Open Date	Payment Status	Time Past Due
Charge Card	\$1,234.56	\$123.45	12/12/11	Current	0
Secured Card	\$500.00	\$50.00	01/13/12	Current	0
Auto Loan	\$10,000.00	\$200.00	03/13/12	Current	0
Mortgage	\$150,000.00	\$1,000.00	06/13/12	Current	0
Total	\$161,734.56	\$1,373.45			

OPEN ACCOUNTS

Account Number	Account Name	Account Type	Current Balance	Minimum Payment	Open Date	Payment Status	Time Past Due
1234 5678	ABC BANK	Charge Card	\$1,234.56	\$123.45	12/12/11	Current	0
9876 5432	DEF BANK	Secured Card	\$500.00	\$50.00	01/13/12	Current	0
1122 3344	GHI BANK	Auto Loan	\$10,000.00	\$200.00	03/13/12	Current	0
5566 7788	JKL BANK	Mortgage	\$150,000.00	\$1,000.00	06/13/12	Current	0

Credit Report - AFTER



A credit report from Experian, Equifax, and TransUnion. A yellow circle highlights the credit scores: Experian 794, Equifax 783, and TransUnion 777. The report includes sections for account information, payment history, and open accounts.

Account Type	Current Balance	Minimum Payment	Open Date	Payment Status	Time Past Due
Charge Card	\$1,234.56	\$123.45	12/12/11	Current	0
Secured Card	\$500.00	\$50.00	01/13/12	Current	0
Auto Loan	\$10,000.00	\$200.00	03/13/12	Current	0
Mortgage	\$150,000.00	\$1,000.00	06/13/12	Current	0
Total	\$161,734.56	\$1,373.45			

OPEN ACCOUNTS

Account Number	Account Name	Account Type	Current Balance	Minimum Payment	Open Date	Payment Status	Time Past Due
1234 5678	ABC BANK	Charge Card	\$1,234.56	\$123.45	12/12/11	Current	0
9876 5432	DEF BANK	Secured Card	\$500.00	\$50.00	01/13/12	Current	0
1122 3344	GHI BANK	Auto Loan	\$10,000.00	\$200.00	03/13/12	Current	0
5566 7788	JKL BANK	Mortgage	\$150,000.00	\$1,000.00	06/13/12	Current	0

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

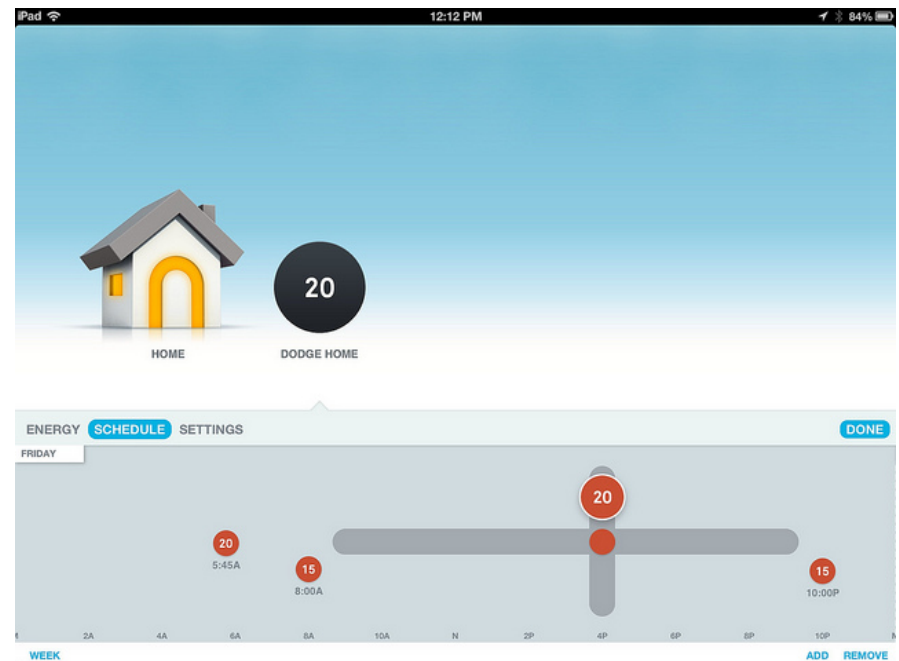
<https://flic.kr/p/6thmfN>

- 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



The screenshot shows the 'SCHEDULE' tab of the Nest Thermostat app. The interface displays a weekly schedule with temperature settings for each day. The temperature is represented by red circles with white numbers. The schedule is as follows:

Day	1	2A	4A	6A	10A	N	2P	4P	6P	8P	1
MONDAY	20			15				20			15
TUESDAY	20			15				20			15
WEDNESDAY	20			15				20			15
THURSDAY	20			15				20			15
FRIDAY	20			15				20			15
SATURDAY				20							15
SUNDAY				20							15

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

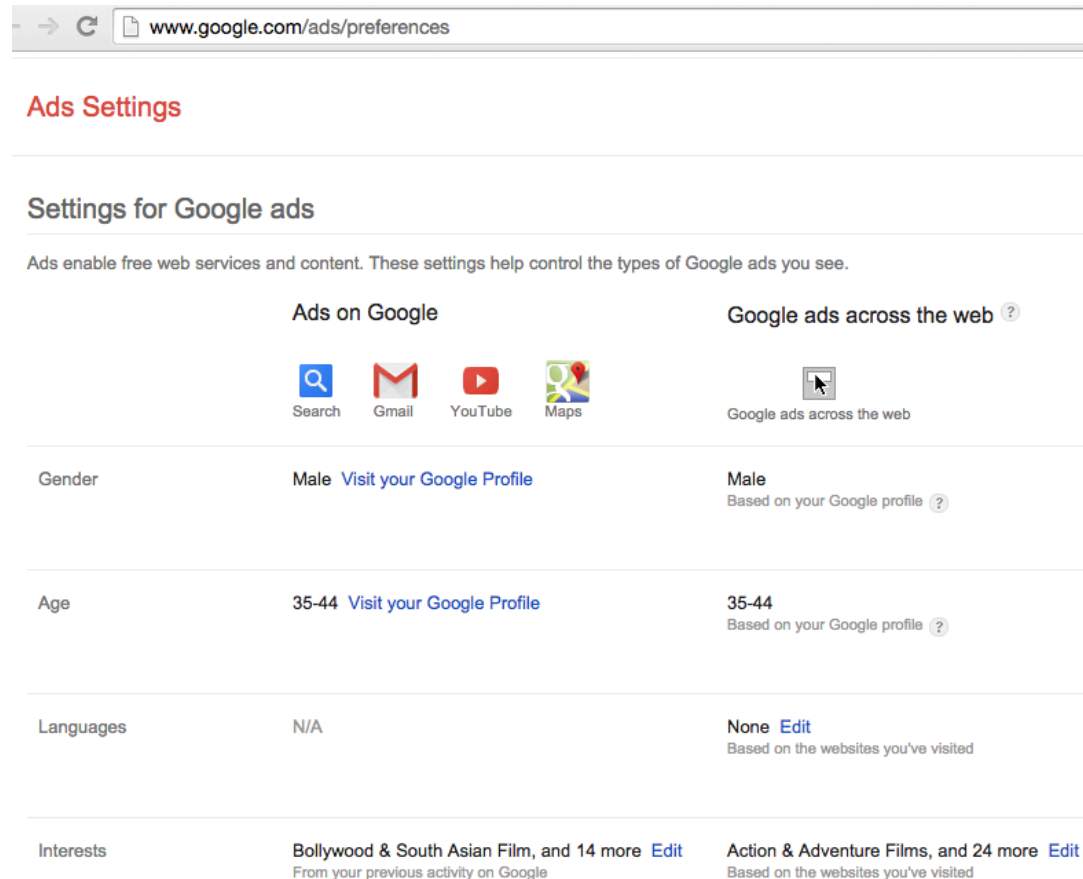
# Lack of Negotiability

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

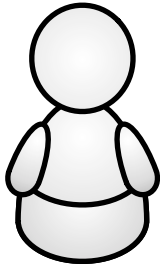


The screenshot shows the Google Ads Settings page at [www.google.com/ads/preferences](https://www.google.com/ads/preferences). The page is titled "Ads Settings" and "Settings for Google ads". It explains that ads enable free web services and content, and these settings help control the types of Google ads you see.

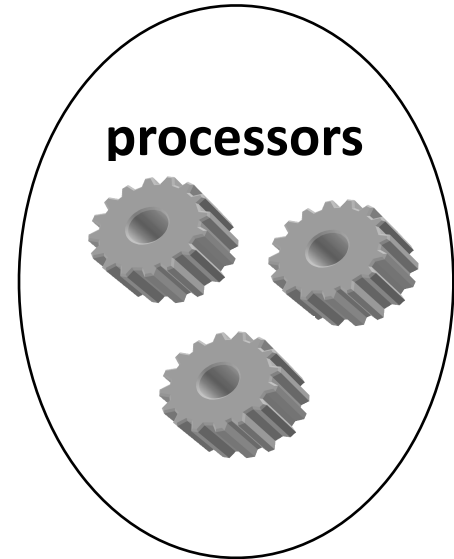
	Ads on Google	Google ads across the web
	 Search  Gmail  YouTube  Maps	 Google ads across the web
Gender	Male <a href="#">Visit your Google Profile</a>	Male Based on your Google profile <a href="#">?</a>
Age	35-44 <a href="#">Visit your Google Profile</a>	35-44 Based on your Google profile <a href="#">?</a>
Languages	N/A	None <a href="#">Edit</a> Based on the websites you've visited
Interests	Bollywood & South Asian Film, and 14 more <a href="#">Edit</a> From your previous activity on Google	Action & Adventure Films, and 24 more <a href="#">Edit</a> Based on the websites you've visited

# Dataware: The Actors

**subject**



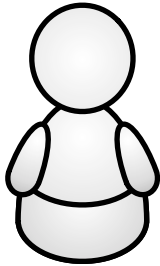
**processors**



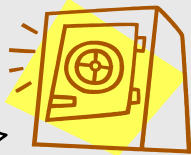


# Dataware: Implementing HDI

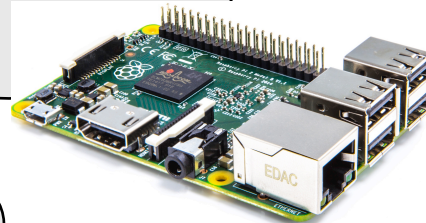
subject



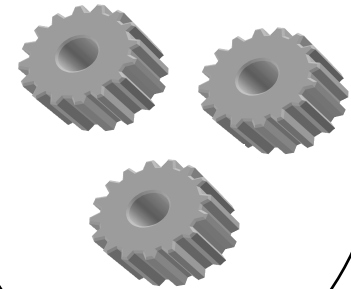
***databox***



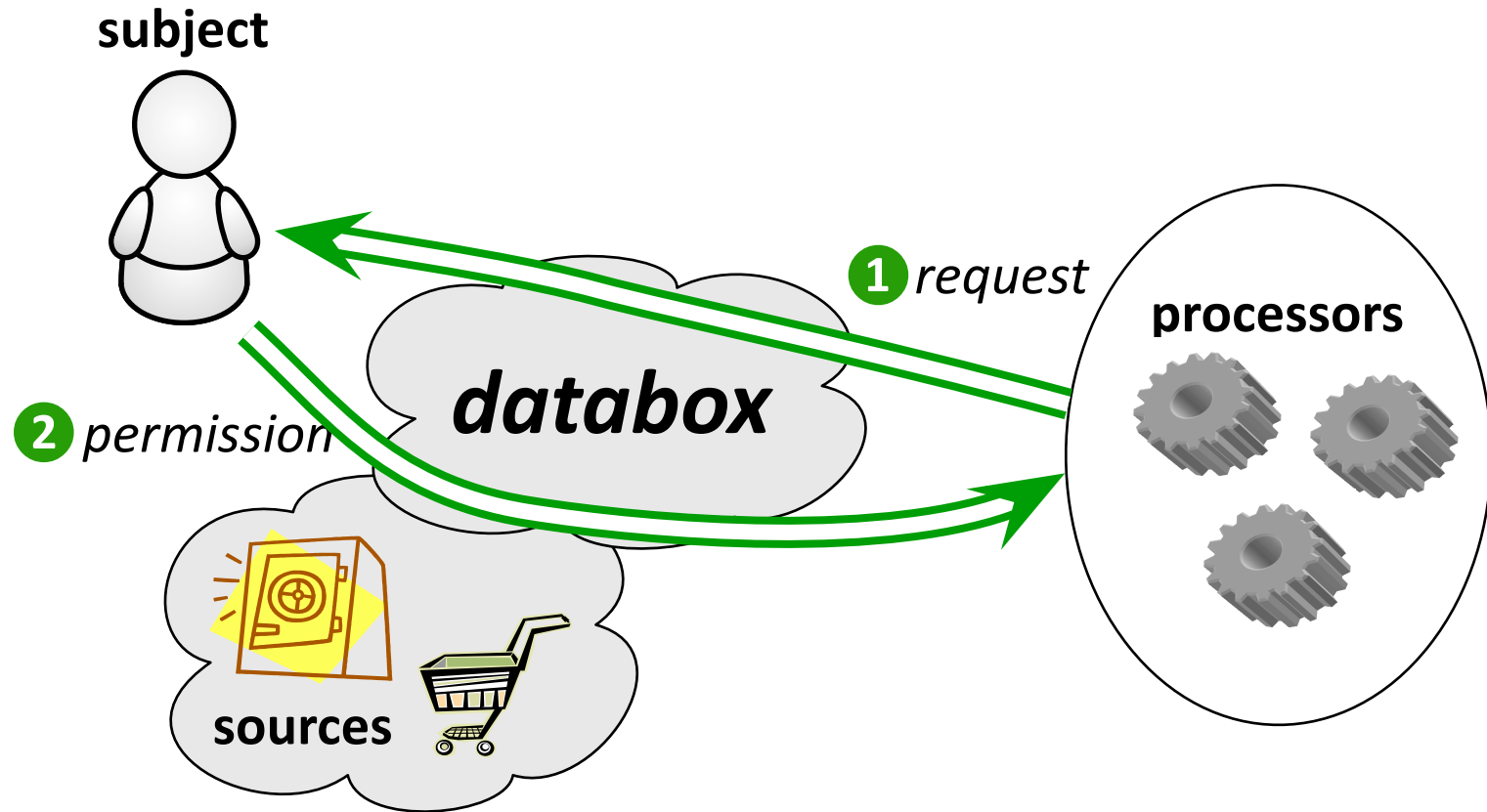
**sources**



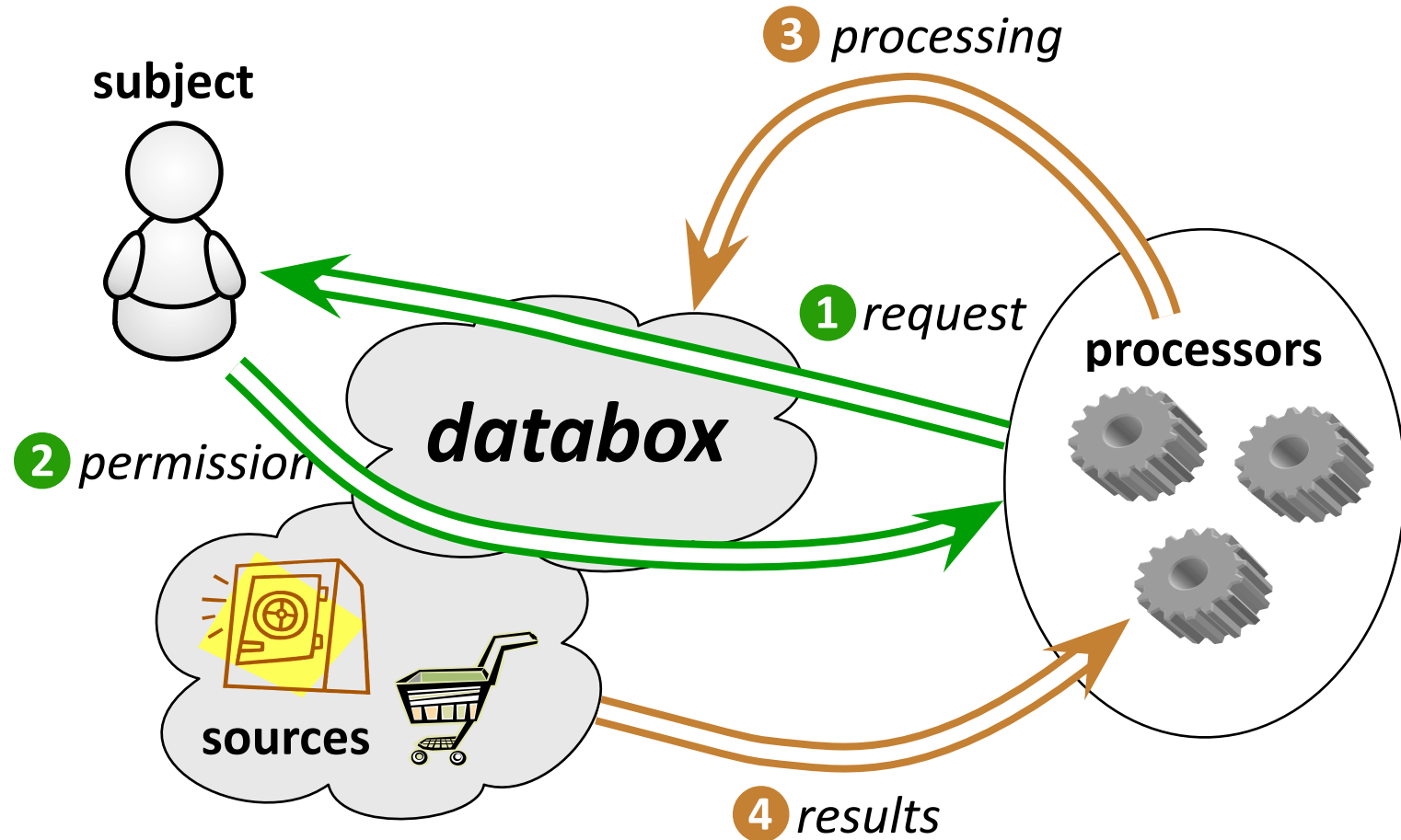
**processors**



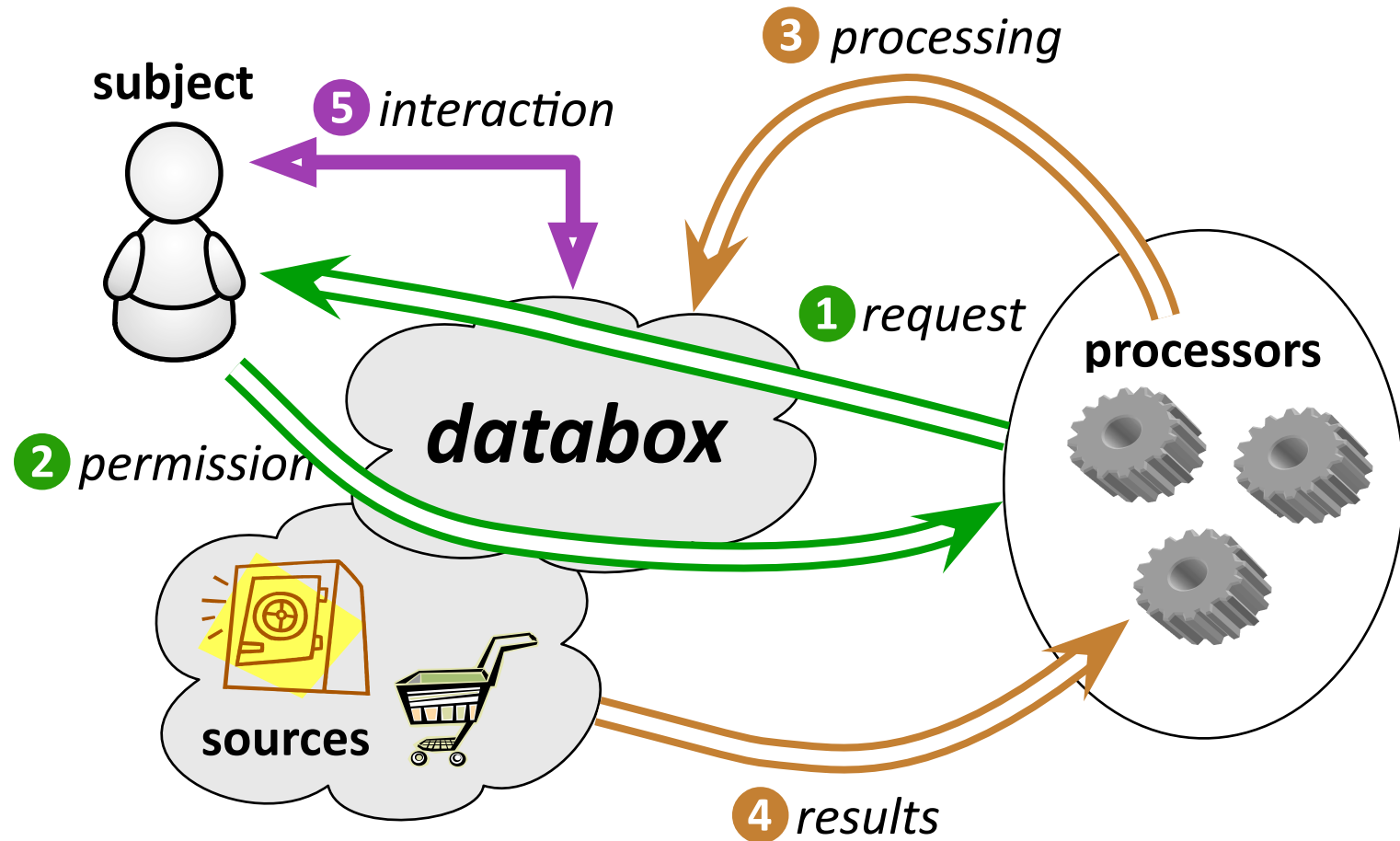
# Dataware: Legibility



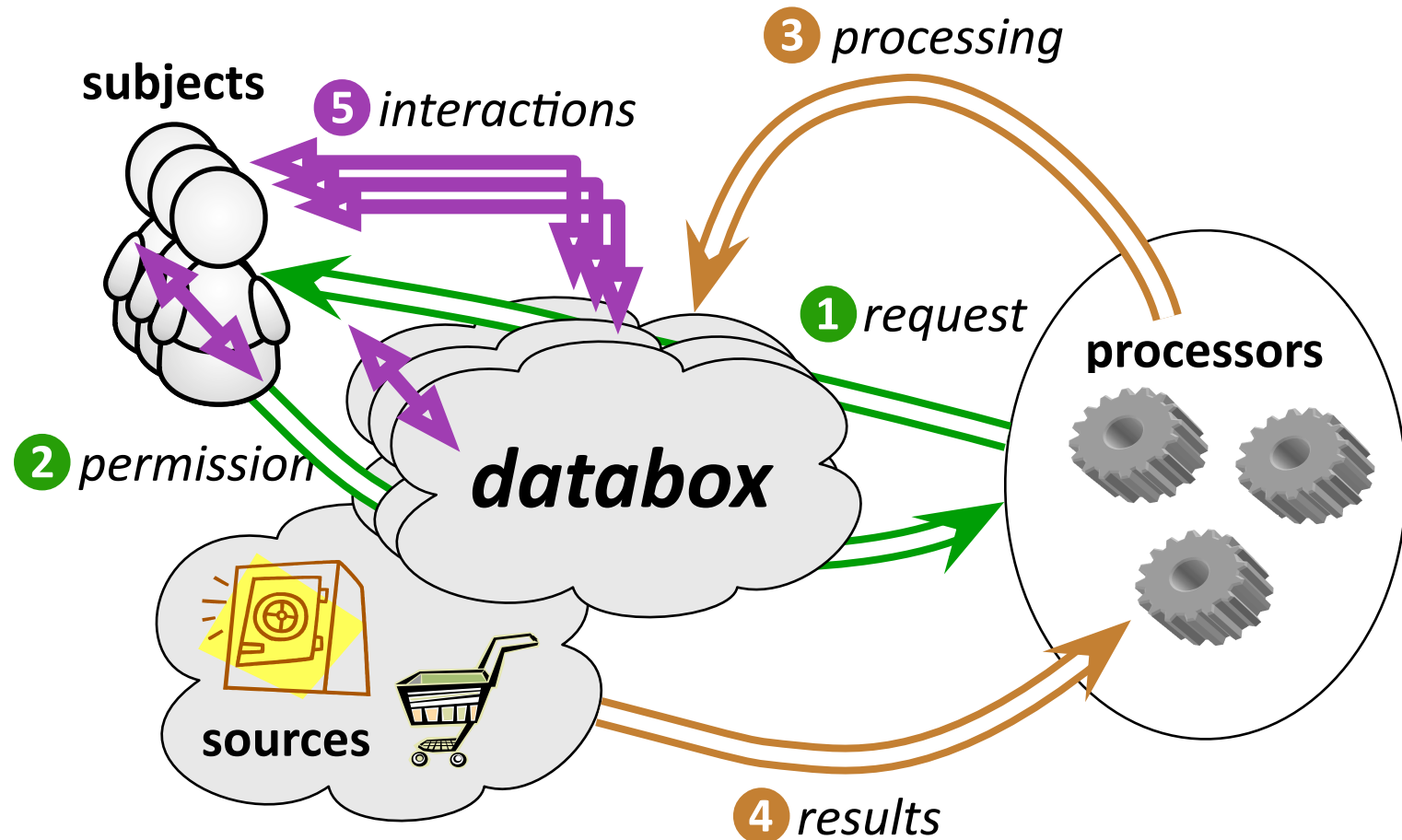
# Dataware: Agency



# Dataware: Negotiability



# Dataware: Constructing Interaction





# Dataware: Constructing Interaction

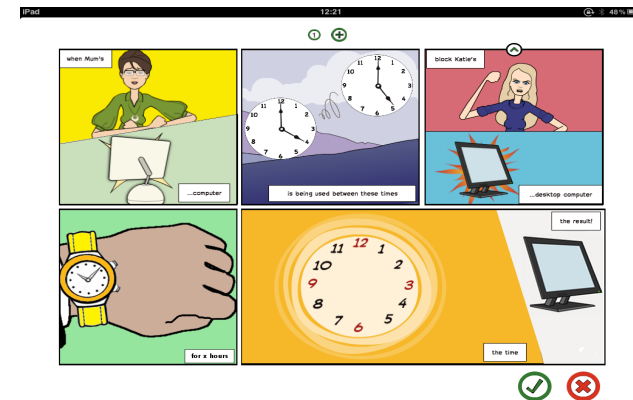
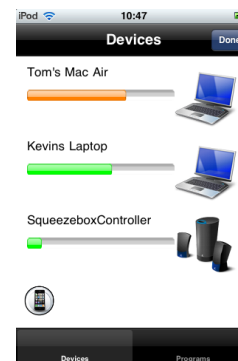
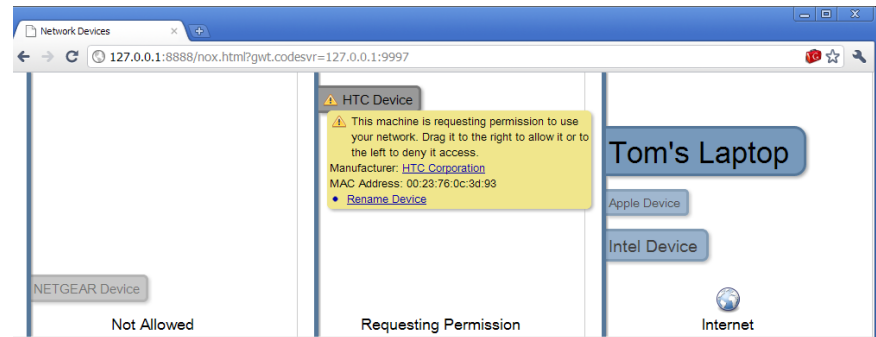
- Numerous proposed interaction models
  - E.g., pay-per-use
- Little about how to actually provide for it
- Dataware one such proposal
  - Accountable transaction between parties in terms of request, permission, audit
- But there's a lot more to consider here...

# 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](mailto: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



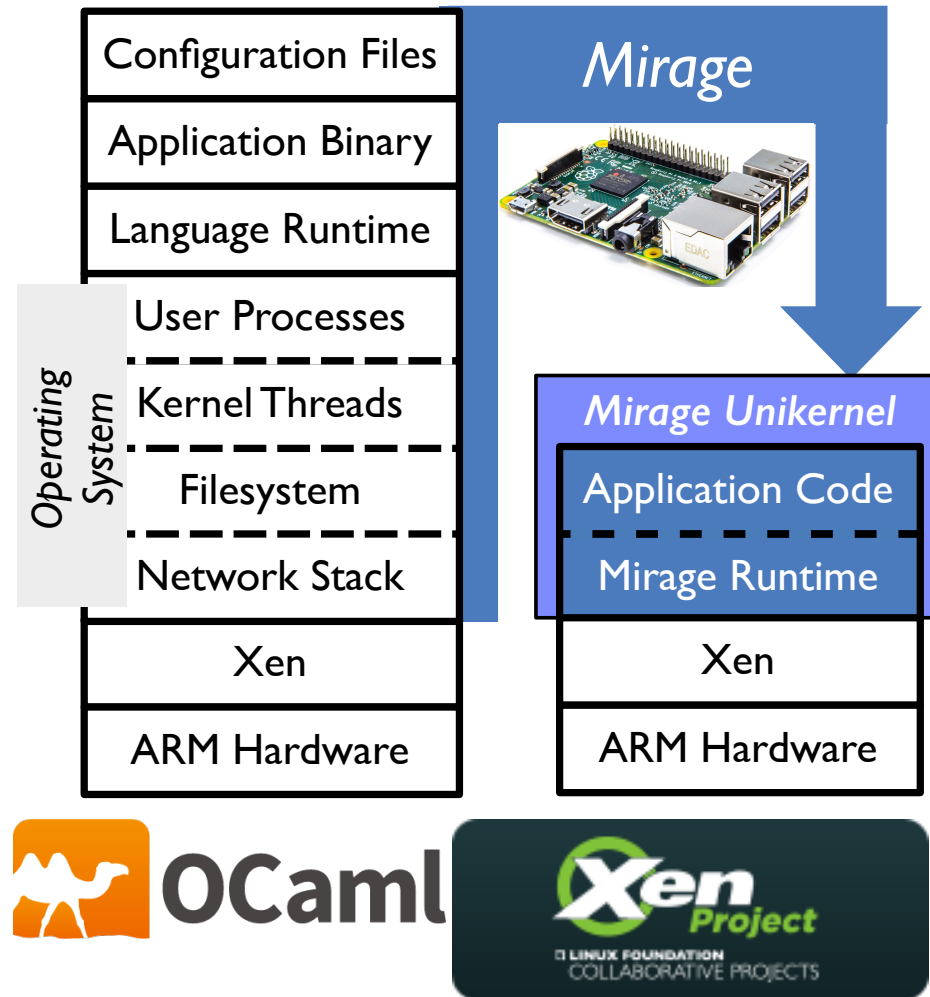
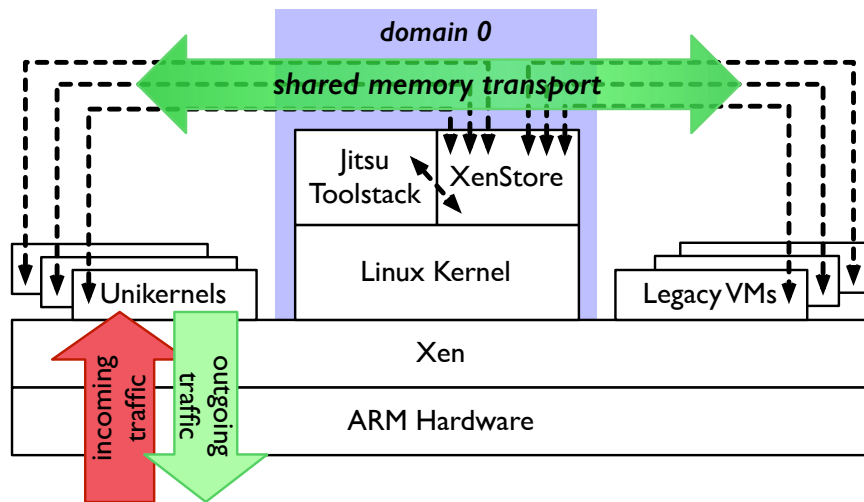
Stable, hidden, shared vs  
Dynamic, exposed, intimate



# 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**

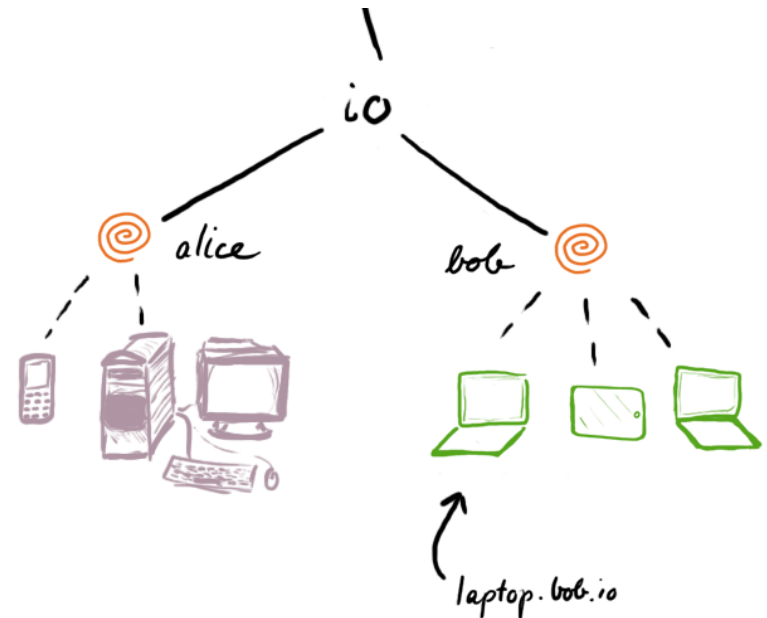
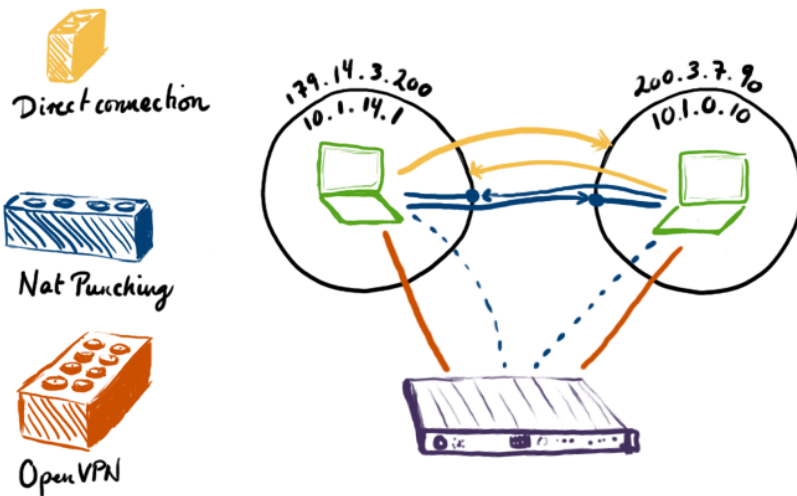




# Signposts

<http://nymote.org/>

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



- Strong identity
- Orchestration
- Security and protection