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

BE DARING
GO TRANSPARENT
OWN YOUR NAME.
MODEL A POSITIVE DIGITAL FOOTPRINT.
SHARE SHAMELESSLY.

In the future, your "digital footprint" will carry far more weight than anything you might include on a resume. -Chris Betcher

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?

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

Inferences, often opaque to users, are drawn from input data and used to drive actions. One action may be to feed inferences back into input data for subsequent analysis.

Actions based on our data and that of others affect our subsequent behaviour.
One action may be to feed inferences back into input data for subsequent analysis.

• 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

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

• E.g., Computation of credit scores

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

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.

https://flic.kr/p/4eQXTn
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
Dataware: The Actors

subject
sources

processors
Dataware: Implementing HDI

Subject:

Databox

Sources:

Processors
Dataware: Legibility

1. request

2. permission

subject

sources

databox

processors
Dataware: Agency

1. request
2. permission
3. processing
4. results

subject

processors

databox

sources
Dataware: Negotiability

1. Request
2. Permission
3. Processing
4. Results
5. Interaction

subject → databox

sources → databox

processors → databox
Dataware: Constructing Interaction

1. Request
2. Permission
3. Processing
4. Results
5. Interactions

- Subjects
- Databox
- Processors
- Sources
- Interaction points
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

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

From My Data to Our Data
- Delegating and revoking control
- Editing, viewing, sharing
- Group management, negotiated collection and control

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

- 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