Personal Control of Your Data

Butler Lampson

August 8, 2013

Background

- What is **new** about online data? It is:
  - **Widespread** in time and space
    - Persistent, easy to copy, visible to anybody
  - **Accessible**: easy to find (by search), connect (by linking)
    - No privacy through obscurity, anonymity is hard
- Data about people in the **physical world** will be just as important as data that is born digital
  - Photos, videos, license plates, location tracks, ...
- Technology and rules must work **hand in hand**
  - Technology **supports** rules, but doesn't determine them
  - "Not allowed to": regulation; "Can't": technology

Principles

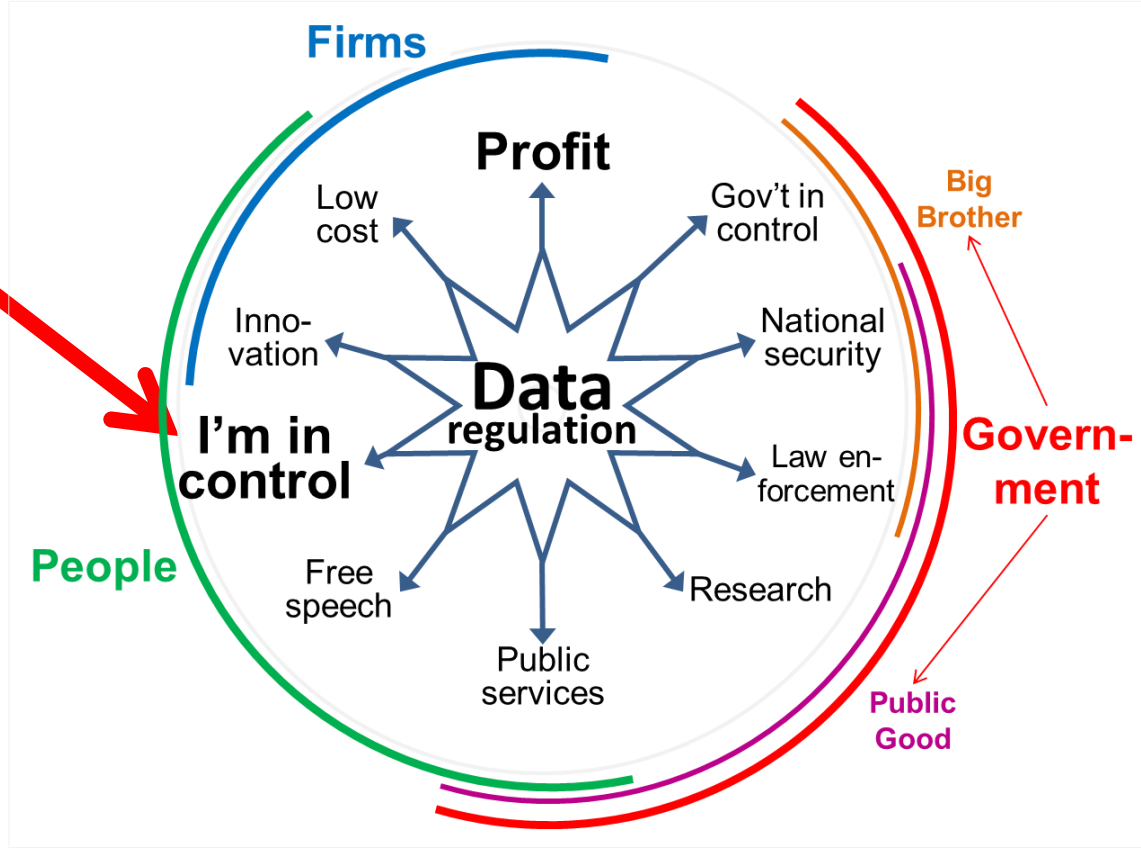
- What is regulation for?
  - To maintain a **balance of power**
    - among people, companies, and governments.
  - To serve the **public good**
    - innovation, research, law enforcement, traffic control, ...
- **Existing law** covers many cases
  - Examples: intellectual property, fraud, public records, ...
- Choices presented to people must be **simple**
  - One screen for the normal case (+ drill-down)
- Regulations **change slowly**, have **unintended consequences**.

More Regulation is Coming

- People: Want **personal control** of their data
  - Even if they know they probably won't exercise it
  - Allow data handlers they trust to access their data
- Regulators: Control of data is a human right
  - Especially the EU, but perhaps US states too
- Firms: Many want consistent, accepted rules, to
  - Build strong relationships with consumers
  - Comply with regulation more easily; safe harbor

Who Wins, Who Loses?

- Regulation serves personal control
- Regulation costs everyone who is regulated



An Ideal for Personal Control

- You keep all your data in a vault you control
- I bring you a query
- If you like the query, you return a result
  - Otherwise you tell me to go away

- This isn't practical
  - Too expensive
  - Too slow
  - Unclear how I may use the result

Practical Personal Control: Goals

- You are empowered to **control** your data
  - **Find** it, limit its **use**, **claim** it
  - **Everywhere**—Across the whole internet
  - **Anytime**, not just when it's collected
  - **Consistently** for all data handlers and devices
  - Remaining **anonymous** if you wish

Practical Personal Control: Mechanisms

- Data tagged with **metadata** that links to policy
- Simple, **coarse-grained policy** and good **defaults**
- **Personas** to manage your different identities
- **No** central database. Instead, two kinds of players:
  - Agents **you choose**—like choosing an email provider
    - **Personal Agent**: handles personas and claiming; can be offline
    - **Policy Service**: tells handlers your policy; must be online
  - **Data handlers**, subject to regulation
    - Anyone who stores or processes your data and is following the rules

Personal Control

- You are empowered to **control** your data:
  - **Find** it, **claim** it
  - Limit its **use**
  - **Anytime**, not just at collection
  - **Everywhere** on the internet
  - **Consistently** for all data handlers and devices
  - With simple, **coarse policy**
    - With good **defaults**
  - **Anonymously** if you wish
    - With **personas** to manage IDs
- **No** central database. Instead
  - Agents **you choose**:
    - Personal agent for personas, claims
    - Policy service to answer handler queries
  - **Data handlers**, regulated

Scenarios

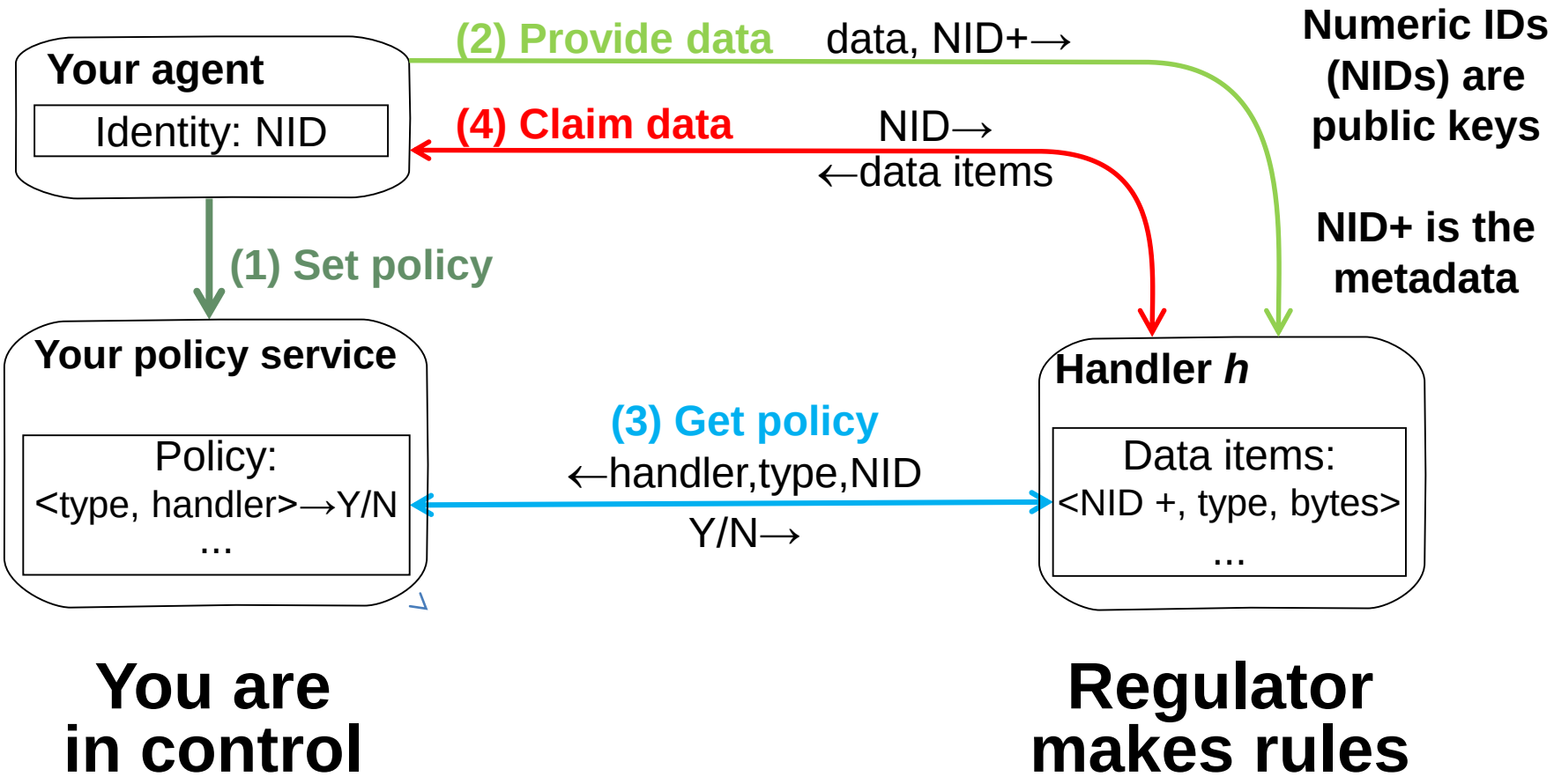
- You move, and you want to know who has your contact information
  - You update some, erase others you don't want
- A school needs to contact a parent in an emergency
  - They use an app that has access to your location data, but reveals only the phone number to call
- You want to see fewer, more interesting ads
  - You disable DoubleClick, keep Neiman-Marcus
- A traffic camera records your license plate
  - DMV records identify you, but you know about the record

How it Works

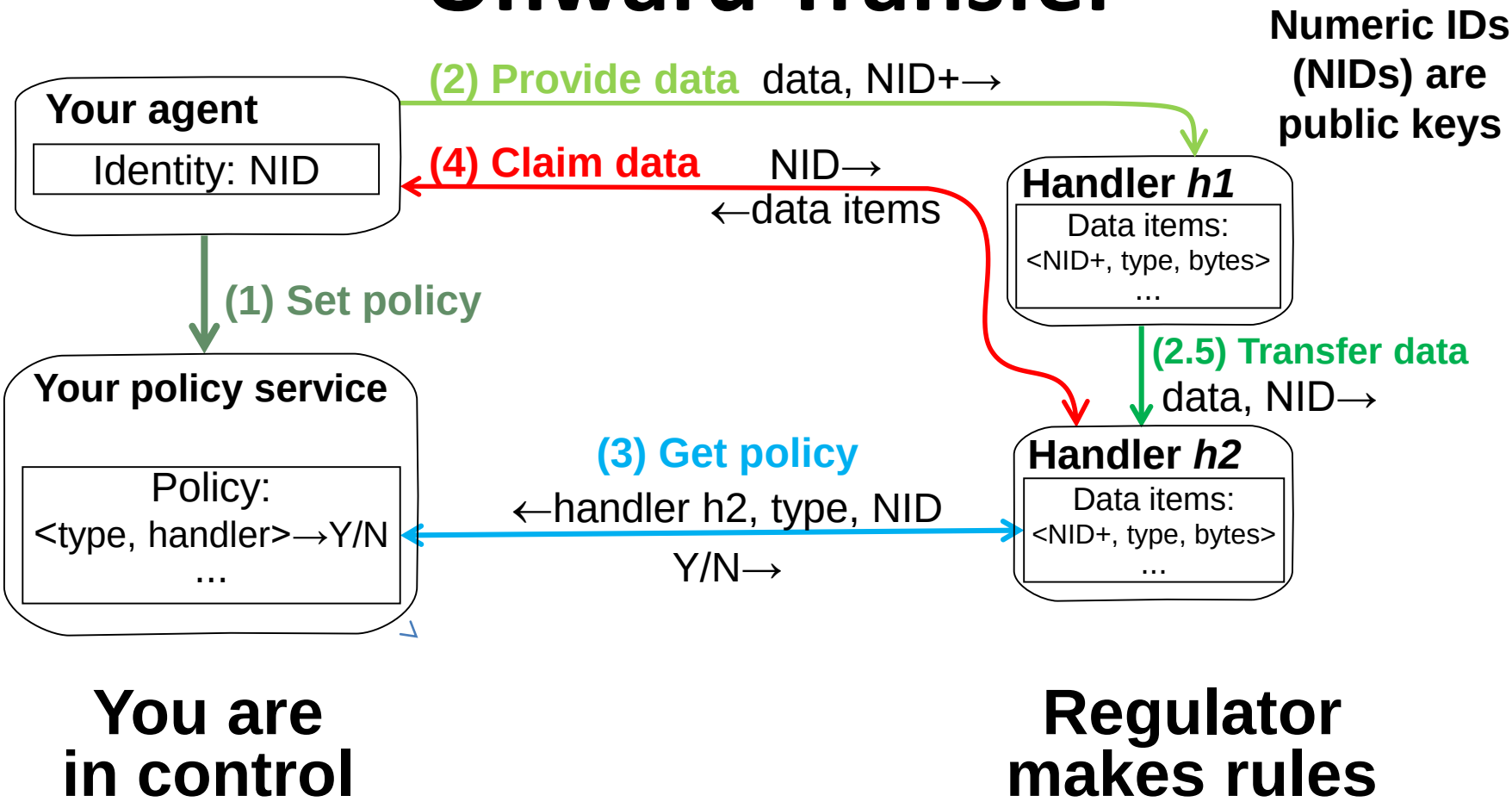
- Data handler **tags** your data with **metadata**
  - Includes a link to your **policy**
  - Your agent supplies it along with your data
  - **Stays** with the data when the data is copied
- Rule: Handler must **check policy** before using data
  - Handler follows policy link and queries **policy service**

- Policy link is NID + URL_{PS}
  - **NID**: Numeric ID
    Anonymized unless you sign in
  - URL_{PS}: to your policy service
- On **re-identification**, handler supplies the metadata
  Especially for **physical world** data—photos, license plates, ...
- Policy service tracks handlers, so people can **find** them
- **Simple** policy, for wide deployment

Who Controls What



Onward Transfer

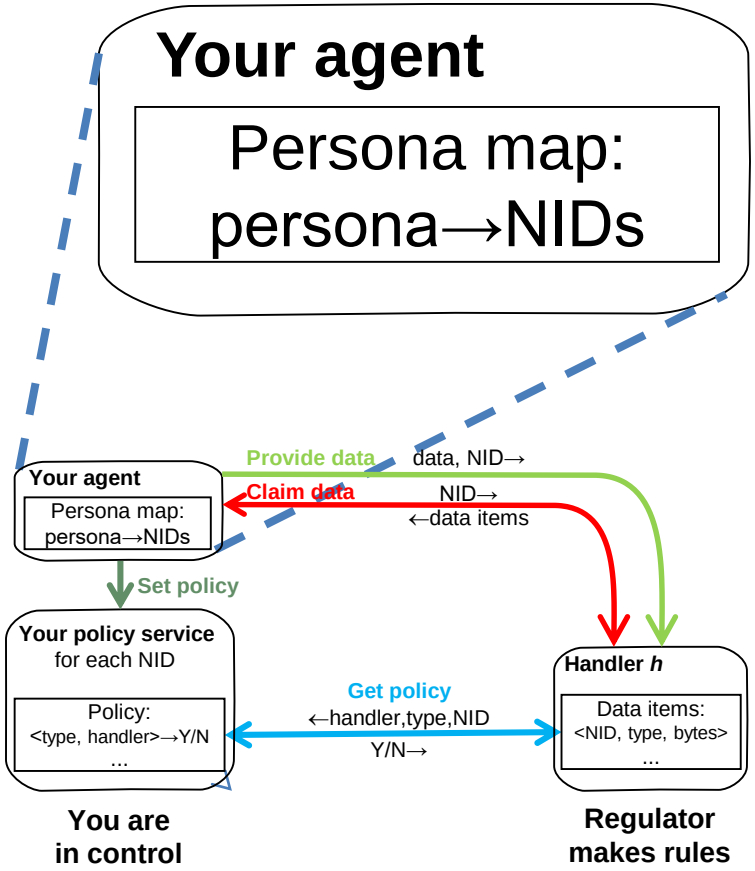


Anonymity

NIDs are public keys
Different relationships call for different kinds of NIDs

Anonymous: Fresh each session
Known: Per web site, tied to cookie
Signed-in: Per account, when signed in

You know about your **personas**
Your persona map tracks <handler, NID>'s used for each persona



Cheaper Anonymous NIDs

NIDs are costly:

- Costly to generate keys
- Costly to store policy for each one

Instead, tag with a **token** that hides NID

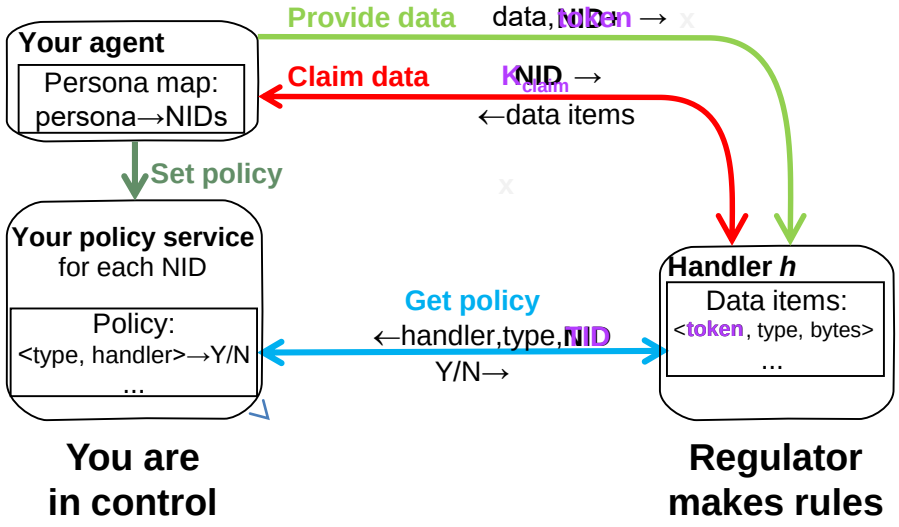
Token = <TID, URL_{PS}, K_{claim}>

- TID = Seal(NID, K_{PS}) different each time
- URL_{PS} points to a popular policy service
- K_{claim} = Hash(TID + K_{person})

TIDs are single-use, so handlers can't link

Policy Service can unseal to get the NID

You can claim data from a handler with K_{claim}



Finding Your Data

Control starts with knowing who has your data
This is tricky:

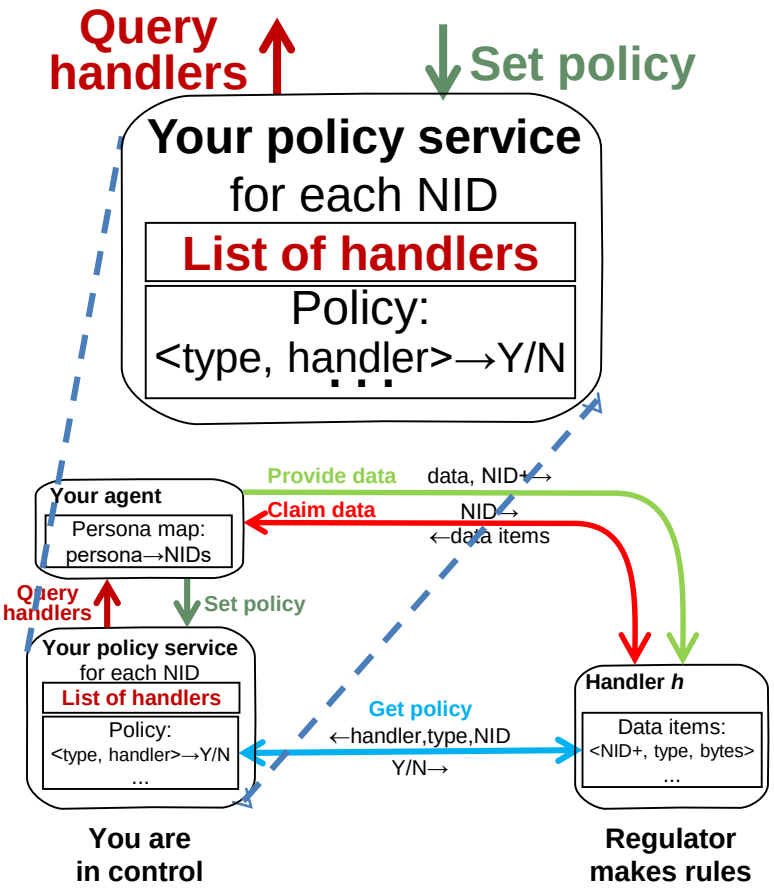
- You talk to **lots** of handlers
- Handlers **transfer** data to other handlers

Policy Service:

- Chosen by you
- Stores policy for each NID
- **Keeps track of handlers**

You can:

- Choose your personas and policy service
- Set policy for your data
- **Query for handlers** that have your data
- Claim your data from a handler



Control vs. Privacy

- There's no free lunch, because of coercion
  - Tracking handlers is useful, but vulnerable
    - Like browsing history
- Forms of coercion
  - Law enforcement/national security
    - Need a warrant or subpoena
  - Personal: parents, spouses, employers, ...
- Mitigations
  - Tell policy service to not track handlers, to delete tracks
  - Transfer tracks to your personal agent
  - Plausible deniability of the true tracks
- Can crypto help?

Policy

- **Data-centric**, not device or service centric
  - Metadata stays with the data, points to the data's policy
- Interface to policy is `<handler, type>`→`Yes/No`
  - Can pass more information, maybe get a richer result
- Basic policy is very simple, for wide deployment
  - 7 ± 2 types of data: contact, location, transaction, ...
    - Can extend a type with a tree of subtypes that can be ignored
  - **Atomic policy**: handler *h* can/can't use data type *t*
  - **Composing** policies: **and**, **or**, **else** on sets of atomic policies
- Encode complex policy in **apps**
  - Treat an app as a handler; the app tags its output suitably

User Experience: Principles

- **One screen** holds most people's policy
  - In big type
  - Drill down to more details, for geeks
- **Templates** (from 3rd parties) + your exceptions
- A reasonable **default** to protect carefree users
  - Easy to change default to a 3rd party template

- Biggest area for future work
  - Only the crudest prototype so far

Refinements

- Metadata stays with data unless it's **aggregated**
  - Need to certify apps that do enough aggregation
- Different personas for personal and **enterprise**
  - The enterprise may manage that persona
- Default for **joint rights**: the parties must agree
  - Agree to allow: Photographer vs. subject
  - Agree to forbid: person vs. public data, e.g., real estate records
- Track **provenance** with extended metadata
  - Log every change, add log pointer to metadata
- **Multiple** policy services, aggregated by your agent
  - Some could be generic, not personal, e.g., Good Housekeeping
- **Extend** policy or data type—ignorable, as in html

Details

- **Changing** your policy service
  - The old one forwards tokens to the new one
  - Optional key escrow for backup
- Control data **uses** through apps
  - Treat an app as a handler, control its access to data
- **Security** of policy queries
  - Handler and policy service authenticate by SSL
- UX for **personas**
  - Make the current persona visible on the screen
  - Default to consistent use of personas on sites

Summary

- More regulation is coming
  - People want **personal control** of their data
- Practical personal control
  - You are empowered to control your data
    - **Find** it, limit its **use**, **claim** it, everywhere, anytime
    - **Consistently** for all data handlers, and **anonymously**
- **Metadata** attached to data, linking to policy
- **Personas** to manage your anonymous identities
- **No** central database