

Personal Control of Your Data

Butler Lampson

Microsoft Research

Computing Research: Addressing
National Priorities and Societal Needs

May 2016

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 links)
 - 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, can't determine them
 - Regulation = not allowed. Technology = impossible

Principles

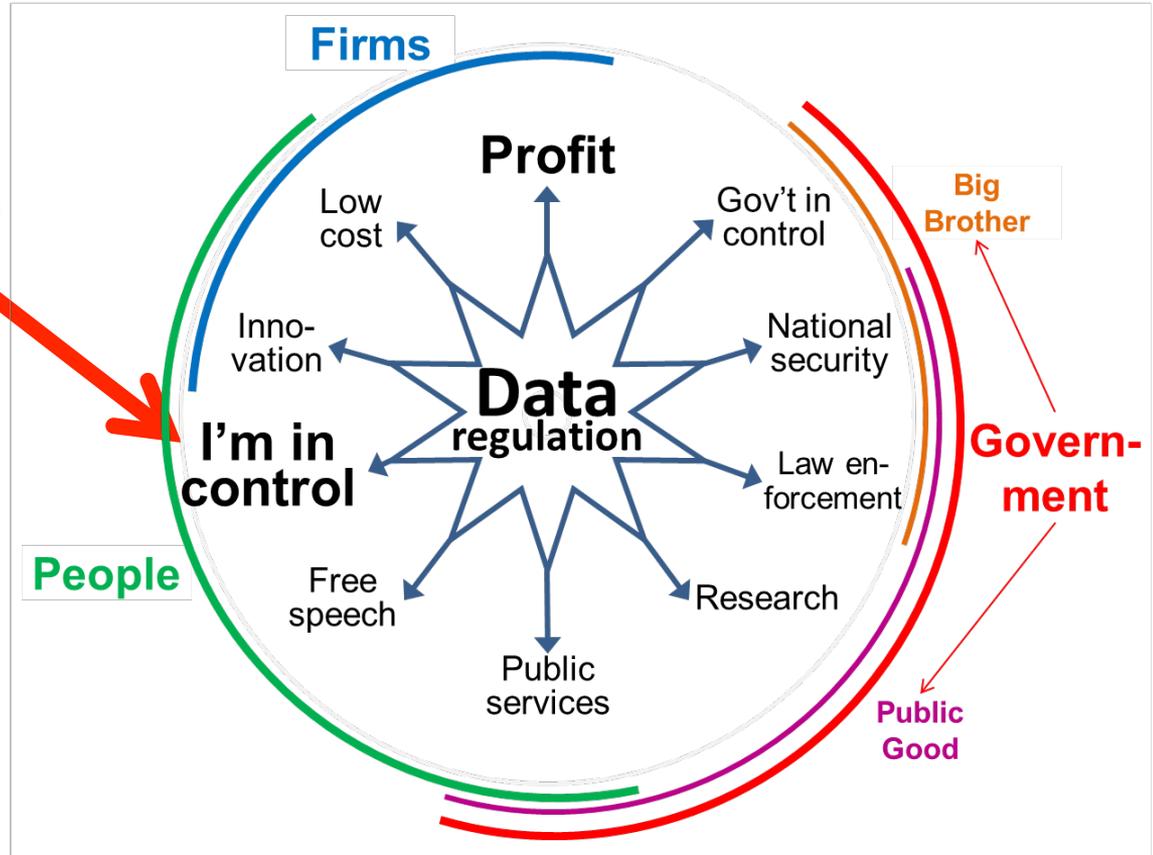
- What is regulation for?
 - To maintain a **balance of power**
 - among people, companies, and governments.
 - To serve the **public good**
 - innovation, research, law enforcement, traffic control,
- **Existing law** covers many cases
 - Examples: intellectual property, fraud, public records, ...
- 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. China?
- 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
 - Prevents aggregating data for research, public good

Practical Personal Control: Goals

- You are empowered to **control** your data
 - **Find** it, **claim** it from handlers
 - Limit its **use**
 - **Anytime**, not just when it's collected
 - **Everywhere**—Across the whole internet
 - **Consistently** for all data handlers and devices
 - With simple, **coarse policy** and good **defaults**
 - Remaining **anonymous** if you wish
 - With **personas** to manage IDs

Practical Personal Control: Mechanisms

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

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

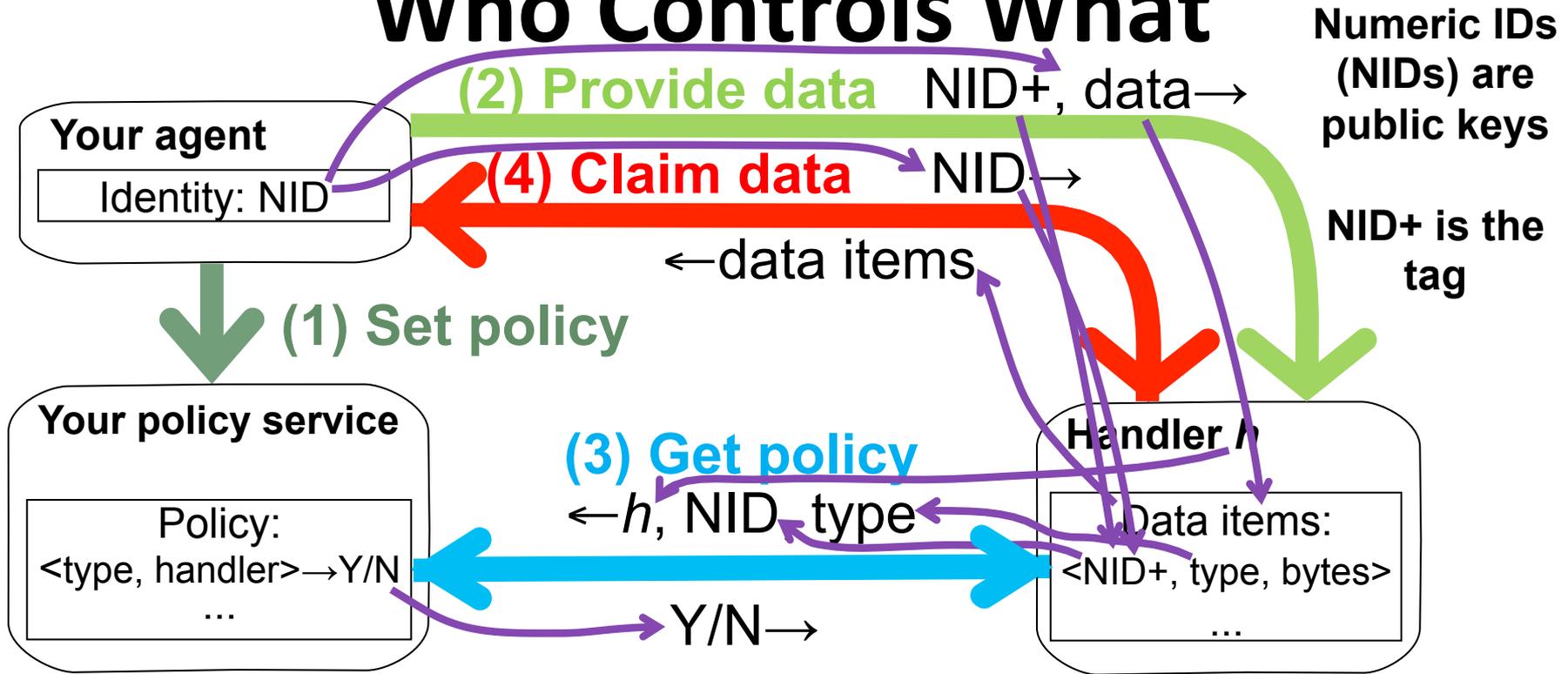
How it Works

- Data handler **tags** your data
 - Includes a link to **your policy**
 - **Comes** with your data from your agent
 - **Stays** with
 - **Copied** data
 - **Computed** results, unless aggregated to preserve privacy
 - **Added** to **re-identified** data
 - Especially for **physical world** data—photos, license plates, ...
- **Rule:** Handler must **check policy** before using data
 - Only works for handlers subject to regulation

Tags

- Tag is $NID + URL_{PS}$
 - **NID**: Numeric ID
 - Anonymized** unless you sign in
 - URL_{PS} : points to your **policy service**
- Handler **queries** policy service through URL_{PS}
- Service **tracks** handlers, so people can find them
- **Simple** policy, for wide deployment

Who Controls What

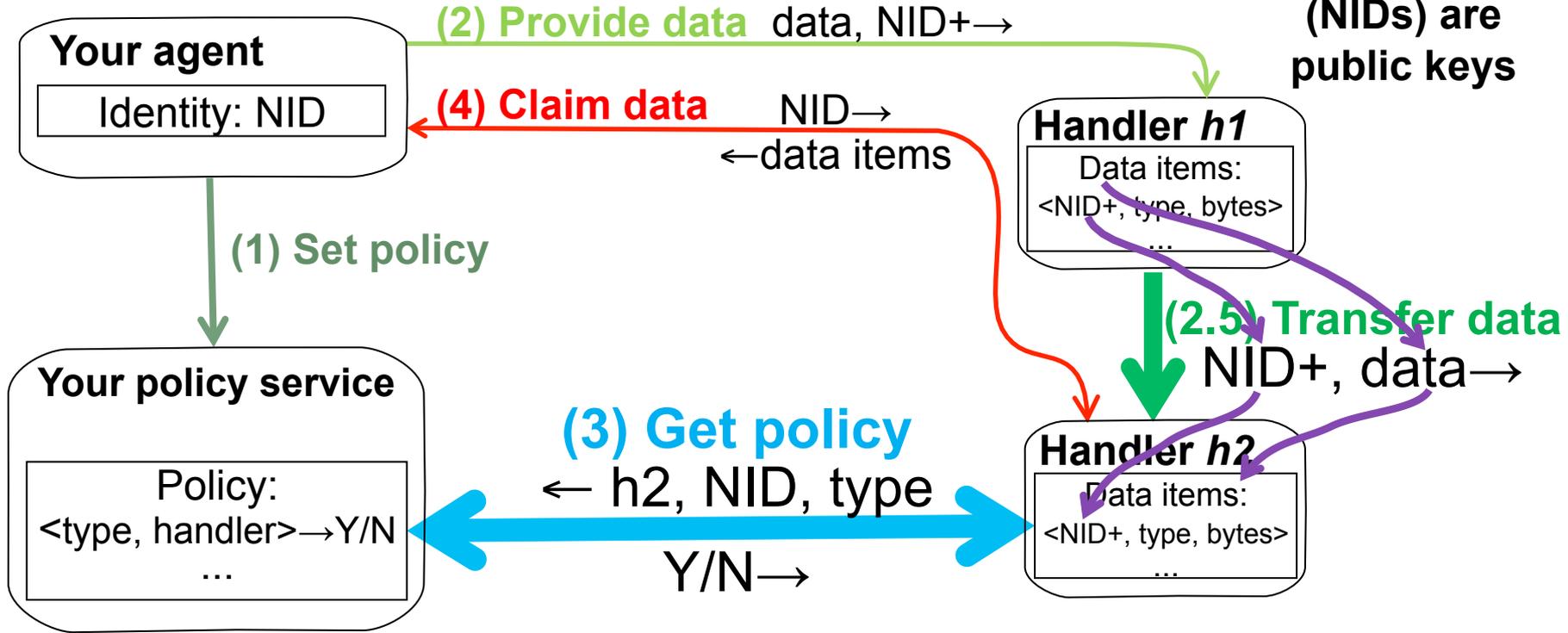


**You are
in control**

**Regulator
makes rules**

Onward Transfer

Numeric IDs
(NIDs) are
public keys



**You are
in control**

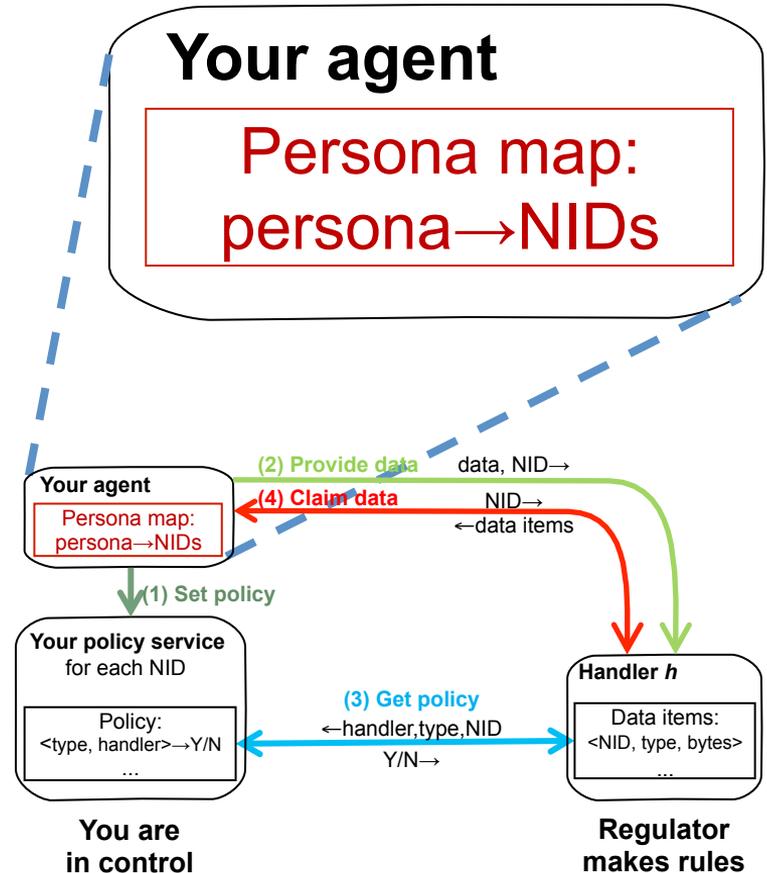
**Regulator
makes rules**

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



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

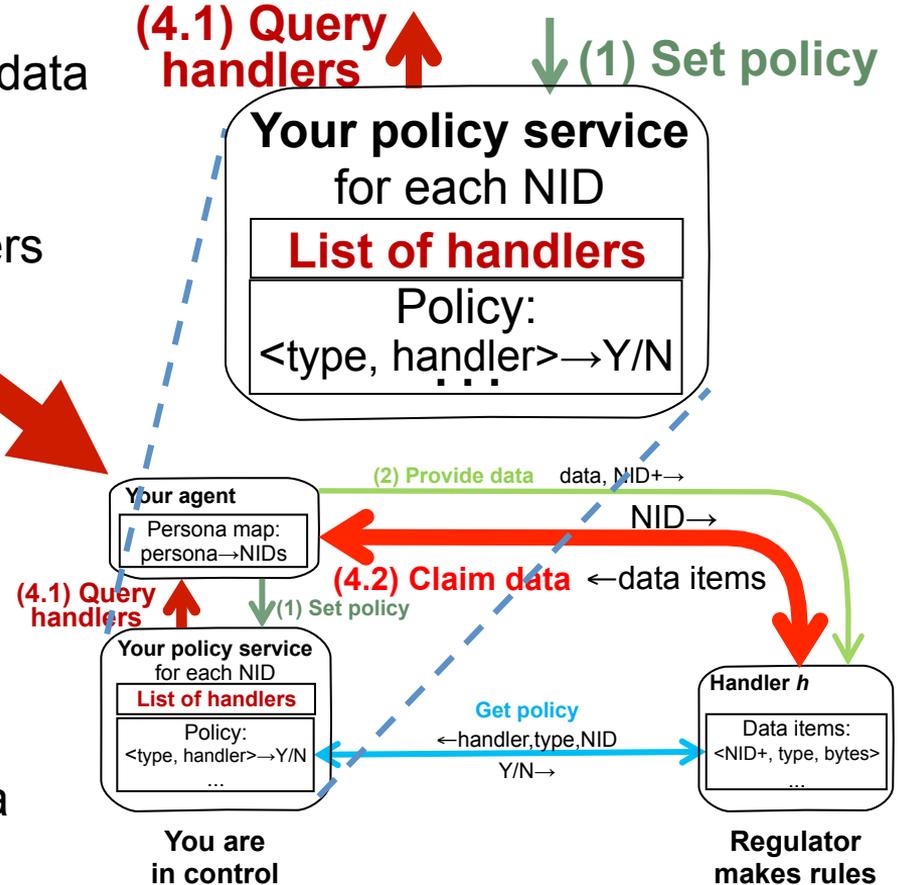
You can:

Choose your personas, 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: people can be **coerced**
 - Tracking handlers is useful, but vulnerable
 - Like browsing history
- **How** to coerce someone
 - Law enforcement/national security
 - Needs authority (warrant, subpoena, ...)
 - Personal: parents, spouses, employers, ...
- **Defending** against coercion
 - Mask true tracks with fakes—plausible deniability

Policy

- **Data-centric**, not device or service centric
 - Tag stays with the data, points to the data's policy
- Simple interface to policy:
 - `<handler, type> → Yes/No`
 - Can pass more information, maybe get a richer result
- Simple standard policy, for wide deployment
 - 7 ± 2 types of data: contact, location, transaction, ...
 - Can extend a type with a tree of subtypes
 - Ignored if not understood

Policy Language

- **Basic policy**
 - handler h may/may not use data type t
- **Composing policies**
 - **and, or, else** on sets of basic policies
- Encode complex policy in **apps**
 - Treat an app as a handler;
 - the app tags its output suitably

User Experience: Principles

- **One screen** shows 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
- This is the biggest area for future work
 - Only the crudest prototype so far

Details

- **Changing** your policy service
 - The old one forwards queries to the new one
 - Optional key escrow for backup
- **Security** of policy queries
 - Handler and policy service authenticate by TLS
- Control complex **uses** of data through apps
 - Treat an app as a handler, control its access to data
- UX for **personas**—make multiple accounts easy
 - Show the current persona; default based on site

Refinements

- Remove tags from data by **aggregation**
 - Need to certify apps that do enough aggregation
- Default for **joint rights**: the parties must agree
 - Agree to allow: Photographer vs. subject
 - Agree to forbid: person vs. public, e.g., real estate records
- Different personas for **personal** and **enterprise**
 - The enterprise may manage its persona
- **Multiple** policy services, managed by your agent
 - Some may be generic, not personal, e.g., Consumer Reports

More Refinements

- **Extend** policy or data type
 - Ignorable, as in html
- Track **provenance** by extending the tag
 - Log every action, add a log pointer to tag

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, **anonymously**
- **Tags** attached to data, linking to policy
- **Personas** to manage your anonymous identities

Personal Control

- You are empowered to **control** your data:
 - **Find** it, **claim** it from handlers
 - 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