P3P and Hippocratic Databases
Privacy Metadata and its Implication for Data Management

Marc Langheinrich
Institute for Pervasive Computing
ETH Zurich
Today’s Program

- **Background: Privacy**
  - Motivation
  - Legal Aspects
- **Platform for Privacy Preferences (P3P)**
  - Goals
  - Specification
- **Hippocratic Databases**
  - Principles
  - Implementation Issues
What is Privacy?

- „The right to be let alone."
  - Louis Brandeis, 1890 (Harvard Law Review)
- “Numerous mechanical devices threaten to make good the prediction that ‘what is whispered in the closet shall be proclaimed from the housetops’”

Louis D. Brandeis, 1856 - 1941
What is Privacy?

The desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude and their behavior to others.

- Alan Westin („Privacy And Freedom“, 1967)
  Prof. Emeritus, Columbia University
Privacy Facts

- **Informational Privacy**
  - Personal Information

- **Communication Privacy**
  - Phone Calls, Letters, E-Mail, ...

- **Territorial Privacy**
  - Privacy of the Home, Office, Car, ...

- **Bodily Privacy**
  - Strip Searches, Drug Testing, ...
Why Privacy?

- Reasons for Privacy
  - Free from Nuisance

Louis D. Brandeis, 1856 – 1941
„The right to be let alone“ (1890)
Why Privacy?

- Reasons for Privacy
  - Free from Nuisance
  - Intimacy

Erving M. Goffman, 1922 – 1982
The Presentation of Self in Everyday Life (1959)
Why Privacy?

- Reasons for Privacy
  - Free from Nuisance
  - Intimacy
  - Free to Decide for Oneself

Beate Rössler

Protecting the decisional autonomy in one’s life (2001)
Why Privacy?

- Reasons for Privacy
  - Free from Nuisance
  - Intimacy
  - Free to Decide for Oneself

- By Another Name...
  - Data Protection
  - Informational Self-Determination

Privacy isn’t just about keeping secrets – data exchange and transparency are key issues!

Beate Rössler

Protecting the decisional autonomy in one’s life (2001)
A (Very) Brief History of Privacy

- Justices Of The Peace Act (England, 1361)
  - Sentences for Eavesdropping and Peeping Toms
- „The poorest man may in his cottage bid defiance to all the force of the crown. It may be frail; its roof may shake; the wind may blow through it; the storms may enter; the rain may enter – but the king of England cannot enter; all his forces dare not cross the threshold of the ruined tenement“
  - William Pitt the Elder (1708-1778)
    English Parliamentarian
    Addressing the House of Commons in 1763
Privacy in the 20th Century

  - “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.“

- 1970 Europäische Menschenrechtskonvention: Artikel 8 – Recht auf Achtung des Privat- und Familienlebens
  - “Everyone has the right to respect for his private and family life, his home and his correspondence.“

- First Data Protection Law in the World in Hesse (1970)
Laws and Regulations

- **Two Main Approaches**
  - Sectorial ("Don’t Fix if it Ain’t Broken")
  - Omnibus (Precautionary Principle)

- **US: Sector-specific Laws, Minimal Protections**
  - Strong Federal Laws for Government
  - Self-Regulation, Case-by-Case for Industry

- **Europe: Omnibus, Strong Privacy Laws**
  - Law Applies to Both Government & Industry
  - Privacy Commissions in Each Country as Watchdog
US Public Sector Privacy Laws

- Federal Communications Act, 1934, 1997 (Wireless)
- Omnibus Crime Control and Safe Street Act, 1968
- Bank Secrecy Act, 1970
- Privacy Act, 1974
- Right to Financial Privacy Act, 1978
- Privacy Protection Act, 1980
- Computer Security Act, 1987
- Family Educational Right to Privacy Act, 1993
- Electronic Communications Privacy Act, 1994
- Driver’s Privacy Protection Act, 1994, 2000
US Private Sector Laws

- Fair Credit Reporting Act, 1971, 1997
- Cable TV Privacy Act, 1984
- Video Privacy Protection Act, 1988
- Health Insurance Portability and Accountability Act, 1996
- Children‘s Online Privacy Protection Act, 1998
- Gramm-Leach-Bliley-Act (Financial Institutions), 1999
EU Data Directive

- 1995 Data Protection Directive 95/46/EC
  - Sets a Benchmark For National Law For Processing Personal Information In Electronic And Manual Files
  - Facilitates Data-flow Between Member States And Restricts Export Of Personal Data To „Unsafe“ Non-EU Countries
  - Follows OECD Fair Information Practices (1980)
    - Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Participation, Accountability
Privacy around the World

- **Australia**
  - Proposed: Privacy Amendment (Private Sector) Bill in 2000
  - In talks with EU officials *(Status?)*

- **Argentina**
  - Passed: Personal Data Protection Act No. 25.326 in 2000
  - EU-certified safe third country

- **Canada**
  - Passed: Bill C-6 in 4/2000
  - EU-certified safe third country

- **Hong Kong**
  - Passed: Personal Data (Privacy) Ordinance in 1995

- **Japan**
  - Traditionally: self-regulation & prefectural laws
  - New law as of 2005 *(Compliance?)*

- **Russia**
  - In Progress: updated to comply with EU directive *(Status?)*

- **South Africa**
  - Planned: Privacy and Data Protection Bill *(Status?)*

- **Switzerland**
  - Data Protection Act of 1992
  - EU-certified safe third country

* Has National Privacy Commissioner

http://www.privacyinternational.org/
Safe Harbor

- **Voluntary Membership**
  - US companies self-certify adherence to requirements
  - Dept. of Commerce maintains list (574 as of 09/04)
    [http://www.export.gov/safeharbor/shoverview.html](http://www.export.gov/safeharbor/shoverview.html)

- **Signatories must provide**
  - **notice** of data collected, purposes, and recipients
  - **choice** of opt-out of 3rd-party transfers, opt-in for sensitive data
  - **access** rights to delete or edit inaccurate information
  - **security** for storage of collected data
  - **enforcement** mechanisms for individual complaints

- **Approved July 26, 2000 by EU**
  - reserves right to renegotiate if remedies for EU citizens prove to be inadequate
EU Data Directive

- **1995 Data Protection Directive 95/46/EC**
  - Sets a Benchmark For National Law For Processing Personal Information In Electronic And Manual Files
  - Facilitates Data-flow Between Member States And Restricts Export Of Personal Data To „Unsafe“ Non-EU Countries
  - Follows OECD Fair Information Practices (1980)
    - Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Participation, Accountability
Fair Information Principles (FIP)

- **Drawn up by the OECD, 1980**
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Participation, Accountability
Fair Information Principles (FIP)

- **Drawn up by the OECD, 1980**
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

- **Five Principles (simplified)**
  1. Openness
  2. Data access and control
  3. Data security
  4. Data minimization
  5. Data subject’s consent

- **Core principles of most modern privacy laws (e.g., Directive)**
  - Implication: Technical privacy solutions must support FIP
**Fair Information Principles (FIP)**

- **Drawn up by the OECD, 1980**
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

- **Five Principles (simplified)**
  1. Openness
  2. Data access and control
  3. Data security
  4. Data minimization
  5. Data subject’s consent

- **Core principles of most modern privacy laws (e.g., Directive)**
  - Implication: Technical privacy solutions must support FIP
Openness (Transparency) on the Web

- **Privacy Policies**
  - Let consumers know about collector’s privacy practices

- **Consumers can then decide**
  - whether or not practices are acceptable
  - when to opt-in or opt-out
  - who to do business with

- **Increase consumer trust**
Privacy Policies Drawbacks

- **But**, Policies are Often...
  - difficult to understand
  - hard to find
  - lengthy to read
    - usually 3-4 pages!
  - changed without notice

Amazon.com Privacy Policy
Technical Solution: P3P
Platform for Privacy Preferences Project (W3C)

- Machine-readable data collection practices (Policy)
  - Who collects and/or processes the data?
  - What information is collected?
  - For what purpose is this data collected?

- Basis-Dataschema
  - Example: user.home.postal.street

- Web-Protocol
  - For exchanging policies between server und browser

P3P Policies

- Machine-readable (XML) version of web site privacy policies
- Use P3P Vocabulary to express data practices
- Use P3P Base Data Set to express type of data collected
- Capture common elements of privacy policies but may not express everything (sites may provide further explanation in human-readable policies)
The P3P Vocabulary

- **Who** is collecting data?
- **What data** is collected?
- For **what purpose** will data be used?
- Is there an ability to **change preferences** about (opt-in or opt-out) of some data uses?
- Who are the data **recipients** (anyone beyond the data collector)?
- To what information does the data collector provide **access**?
- What is the data **retention** policy?
- How will **disputes** about the policy be resolved?
- Where is the **human-readable privacy policy**?
P3P Base Data Schema

- A set of common data elements that all P3P implementations should know about
- Includes user, thirdparty, and business elements such as name, address, phone number, etc.
P3P Base Data Schema

- A set of common data elements that all P3P implementations should know about.
- Includes *user, thirdparty, and business* elements such as name, address, phone, etc.

```
business.contact-info.telecom.fax.ext
```
P3P Base Data Schema

- A set of common data elements that all P3P implementations should know about
- Includes user, thirdparty, and business elements such as name, address, phone number, etc.
- Includes “Dynamic” elements such as indicators that a site collects click-stream, uses cookies, collects info of a certain category, etc.
Example Privacy Policy

At CatalogExample, we care about your privacy. When you come to our site to look for an item, we will only use this information to improve our site and will not store it in an identifiable way.

CatalogExample is a licensee of the PrivacySealExample Program. ...

Questions regarding this statement should be directed to: CatalogExample 1-248-392-6753

When you browse through our site we collect:

- The basic information about your computer and connection to make sure that we can get you the proper information and for security purposes
- Aggregate information on what pages consumers access or visit to improve our site

We purge the browsing information that we collect regularly
P3P/XML Encoding

<POLICY xmlns="http://www.w3.org/2000/12/P3Pv1"
    discuri="http://www.catalog.example.com/Privacy.html">

    <ENTITY>
        <DATA-GROUP>
            <DATA ref="#business.name">CatalogExample</DATA>
            <DATA ref="#business.tel.intcode">1</DATA>
            <DATA ref="#business.tel.loccode">248</DATA>
            <DATA ref="#business.tel.number">3926753</DATA>
        </DATA-GROUP>
    </ENTITY>

    <ACCESS><nonident/></ACCESS>

    <DISPUTES-GROUP>
        <DISPUTES resolution-type="independent">
            service="http://www.PrivacySeal.example.org"
            short-description="PrivacySeal.exampleorg">
                <REMEDIES><correct/></REMEDIES>
                <IMG src="http://www.PrivacySeal.example.org/Logo.gif"/>
        </DISPUTES>
    </DISPUTES-GROUP>

    <STATEMENT>
        <PURPOSE><admin/><develop/></PURPOSE>
        <RECIPIENT><ours/></RECIPIENT>
        <DATA-GROUP>
            <DATA ref="#dynamic.clickstream"/>
            <DATA ref="#dynamic.http"/>
        </DATA-GROUP>
        <DATA-GROUP>
            <RETENTION><stated-purpose/></RETENTION>
        </DATA-GROUP>
    </STATEMENT>
</POLICY>
Technical Solution: P3P
Platform for Privacy Preferences Project (W3C)

- Machine-readable data collection practices (Policy)
  - Who collects and/or processes the data?
  - What information is collected?
  - For what purpose is this data collected?

- Basis-Dataschema
  - Example: `user.home.postal.street`

- Web-Protocol
  - For exchanging policies between server and browser

Technical Solution: P3P
Platform for Privacy Preferences Project (W3C)

- Machine-readable data collection practices (Policy)
  - Who collects and/or processes the data?
  - What information is collected?
  - For what purpose is the data collected?

- Basis-Dataschema
  - Example: `user.home.postal.street`

- Web-Protocol
  - For exchanging policies between server and browser

Fair Information Principles (FIP)

- Drawn up by the OECD, 1980
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

- Five Principles (simplified)
  1. Openness
  2. Data access and control
  3. Data security
  4. Data minimization
  5. Data subject’s consent

- Core principles of most modern privacy laws
  - Implication: Technical solutions must support FIP
Fair Information Principles (FIP)

- Drawn up by the OECD, 1980
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

- Five Principles (simplified)
  1. Openness
  2. Data access and control
  3. Data security
  4. Data minimization
  5. Data subject’s consent

How to keep your promises?
Fair Information Principles (FIP)

- Drawn up by the OECD, 1980
  - “Organisation for economic cooperation and development”
  - Voluntary guidelines for member states
  - Goal: ease transborder flow of goods (and information)

- Five Principles (simplified)
  1. Openness
  2. Data access and control
  3. Data security
  4. Data minimization
  5. Data subject’s consent

How to keep your promises?

-> Let the database do the work for you
Privacy-Aware Databases: Basic Principle

- All Data is Stored Together With Privacy (P3P) Policy
  - Data and policy (Metadata) form logical unit
- Each Data Access Needs Usage Policy
  - Database compares allowed/announced and proposed usage
  - Data with non-matching allowed usage is held back
  - Each data access (who, why) is recorded (auditing)
Privacy-Aware Databases: Basic Principle

- Database compares allowed/announced usage.
- Data with non-matching allowed usage is held back.
- Each data access (who, why) is recorded (auditing).

Example result of a policy-based query:

```
<table>
<thead>
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>Gender</th>
<th>Email</th>
<th>Street</th>
<th>Zip</th>
<th>City</th>
<th>State</th>
<th>Phone</th>
<th>BirthYear</th>
<th>BirthMonth</th>
<th>BirthDay</th>
<th>LastSeen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack</td>
<td>Doe</td>
<td>m</td>
<td><a href="mailto:jack@example.org">jack@example.org</a></td>
<td>1000 Main St</td>
<td>10234</td>
<td>New York</td>
<td>NY</td>
<td>(206) 342-2939</td>
<td>1976</td>
<td>March</td>
<td>01.01.2005 12:15:23</td>
<td></td>
</tr>
<tr>
<td>Jane</td>
<td>Doe</td>
<td>w</td>
<td><a href="mailto:jane@example.org">jane@example.org</a></td>
<td>1000 Main St</td>
<td>10234</td>
<td>New York</td>
<td>NY</td>
<td>(206) 342-2939</td>
<td>1976</td>
<td>March</td>
<td>01.01.2005 12:15:23</td>
<td></td>
</tr>
</tbody>
</table>
```
Privacy-Aware DBs: “Hippocratic Databases”
Agrawal, Kiernan, Srikant, and Xu (VLDB 2002)

- Inspired by Hippocratic Oath
  - “And about whatever I may see or hear in treatment, or even without treatment, in the life of human beings – things that should not ever be blurted outside – I will remain silent, holding such things to be unutterable”
    - Hippocratic Oath #8
Hippocratic Database Principles
Agrawal, Kiernan, Srikant, and Xu (VLDB 2002)

- Following the Fair Information Principles:
  1. Purpose Specification
  2. User Consent
  3. Limited Collection
  4. Limited Use
  5. Limited Disclosure
  6. Limited Retention
  7. Accuracy
  8. Safety
  9. Openness
  10. Compliance
Hippocratic Databases: Strawman Design
Agrawal, Kiernan, Srikant, and Xu (VLDB 2002)

1. Purpose Specification
2. User Consent
3. Limited Collection
4. Limited Use
5. Limited Disclosure
6. Limited Retention
7. Accuracy
8. Safety
9. Openness
10. Compliance
Privacy Metadata (Controls Disclosure)

- **Purpose Only** (Agrawal et al., VLDB 2002)
  - Probably efficient (row-level limited disclosure)
  - Limited expressability

- **Full P3P Policies** (Langheinrich, Ubicomp 2002)
  - Expressive (cell-level limited disclosure)
  - Complex access control implementation

- **Purpose/Recipient Pairs** (LeFevre et al., VLDB 2004)
  - Sufficiently expressive?
  - Efficient!
**Limited Disclosure Mechanism (Cell-Level)**
LeFevre, Agrawal, Ercegovac, Ramakrishnan, Xu, and DeWitt (VLDB’04)

- **Table Semantics Model**
  - Each purpose-recipient pair defines a table view
  - Views combine to define coherent relational database
  - Queries operate directly on corresponding views

- **Query Semantics Model**
  - Table-references in queries are replaced with condition
  - Faster (fewer records)

---

**Figure 2:** Full data table of patient information.

<table>
<thead>
<tr>
<th>P#</th>
<th>Name</th>
<th>Age</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alice Adams</td>
<td>10</td>
<td>1 April Ave.</td>
<td>111-1111</td>
</tr>
<tr>
<td>2</td>
<td>Bob Blaney</td>
<td>20</td>
<td>2 Brooks Blvd.</td>
<td>222-2222</td>
</tr>
<tr>
<td>3</td>
<td>Carl Carson</td>
<td>30</td>
<td>3 Cricket Ct.</td>
<td>333-3333</td>
</tr>
<tr>
<td>4</td>
<td>David Daniels</td>
<td>40</td>
<td>4 Dogwood Dr.</td>
<td>444-4444</td>
</tr>
</tbody>
</table>

**Figure 3:** Patient choices for disclosure of information to charities for solicitation.

<table>
<thead>
<tr>
<th>P#</th>
<th>P#</th>
<th>Name</th>
<th>Age</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>x</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

**Figure 4:** Privacy-enforced table of patient information, using table semantics.

<table>
<thead>
<tr>
<th>P#</th>
<th>Name</th>
<th>Age</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alice Adams</td>
<td>10</td>
<td>1 April Ave.</td>
<td>111-1111</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>33</td>
<td>3 Cricket Ct.</td>
<td>333-3333</td>
</tr>
<tr>
<td>4</td>
<td>David Daniels</td>
<td></td>
<td>4 Dogwood Dr.</td>
<td>-</td>
</tr>
</tbody>
</table>
Implementation Issues
LeFevre, Agrawal, Ercegovac, Ramakrishnan, Xu, and DeWitt (VLDB’04)

- **Representing Prohibited Values**
  - NULL value well-known (but carries semantic anomalies)
  - e.g. “PROHIBITED” requires new SQL semantics in apps

- **On-the-fly Query Rewriting (Limiting Disclosure)**
  - CASE statements typically faster
  - OUTER JOIN statements better for “sparse” consent

- **Choice (Opt-In, Opt-Out) Storage**
  - Internal (additional columns in table) typically faster
  - External (separate choice table) better for many choices
Performance (Query Processing Overhead)
LeFevre, Agrawal, Ercegovac, Ramakrishnan, Xu, and DeWitt (VLDB’04)

Choice selectivity 100%

Lower Bar shows CPU time, upper bar I/O time.
Impact of Record Filtering
LeFevre, Agrawal, Ercegovac, Ramakrishnan, Xu, and DeWitt (VLDB’04)
Challenges

- Policies on non-read operations?
  - Performance impact on UPDATE, DELETE, etc ops?
- Limited Retention?
  - Impact on DB integrity when deleting logs, checkpoints
- Languages?
  - P3P vs. purpose only vs. purpose-recipient granularity
- Versioning?
  - How to handle policy updates at the DB level?
- Data Export/Import (Intra/Inter-Business)?
Platform for Enterprise Privacy Practices
Karjoth, Schunter, Waidner (PET-Workshop, 2002)

- In-Company Enforcement Methodology for P3P
  - Higher level of policy detail (internal-use only)
  - More complex policy options (e.g., depending on data)

- Sticky Policy Paradigm
  - Data grouped into “Forms” with a single “Type”
  - Form always associated with policies of matching types

- Introduces “Obligations”
  - Gaining data access might incur obligations (i.e., “Data must be deleted if not consented after 7 days”)
Commercial Implementations

- Lucrative Market?
  - IBM’s “Tivoli Privacy Manager”
  - Synomos’ (formerly ZeroKnowledge) “Align”
  - Voltage’s “Enterprise Privacy Management Platform”
  - NCR Teradata

- Exact Semantics and Capabilities Unclear
  - Few freely available information
Additional Uses?

- **Privacy Metadata is Usage Metadata**
  - Not just for privacy purposes: intellectual property, digital content, etc.
  - Implementations often policy-neutral

- **Requires Usage Model (Rights/Threats/Obligations)**
  - How much security, enforcement, auditing is needed?
  - More levels of control increase complexity, limit efficiency
  - Support for billing?
Summary

- Privacy is an Important Social and Legal Issue
  - Solid history
  - Often strong legal protection (especially in Europe)

- Privacy Policies
  - Basis for Fair Information Principles
  - P3P or EPAL for automated processing

- Hippocratic Databases for “Keeping Promises”
  - Foundation for enterprise-wide compliant data flows
  - Implementation non-trivial (promising research!)
  - Also useful for intellectual property, digital goods, ...
Privacy Reads

- **Lawrence Lessig**: *Code and Other Laws of Cyberspace*. Basic Books, 2000
Sample Conferences

- **Computer, Freedom, Privacy (CFP)**
  - April/May (?) 2006, Washington D.C.

- **Very Large Databases (VLDB)**
  - September 2006, Seoul, Korea

- **Secure Data Management Workshop (SDM)**
  - Held at VLDB 2004 and VLDB 2005
  - 3rd installation planned for VLDB 2006