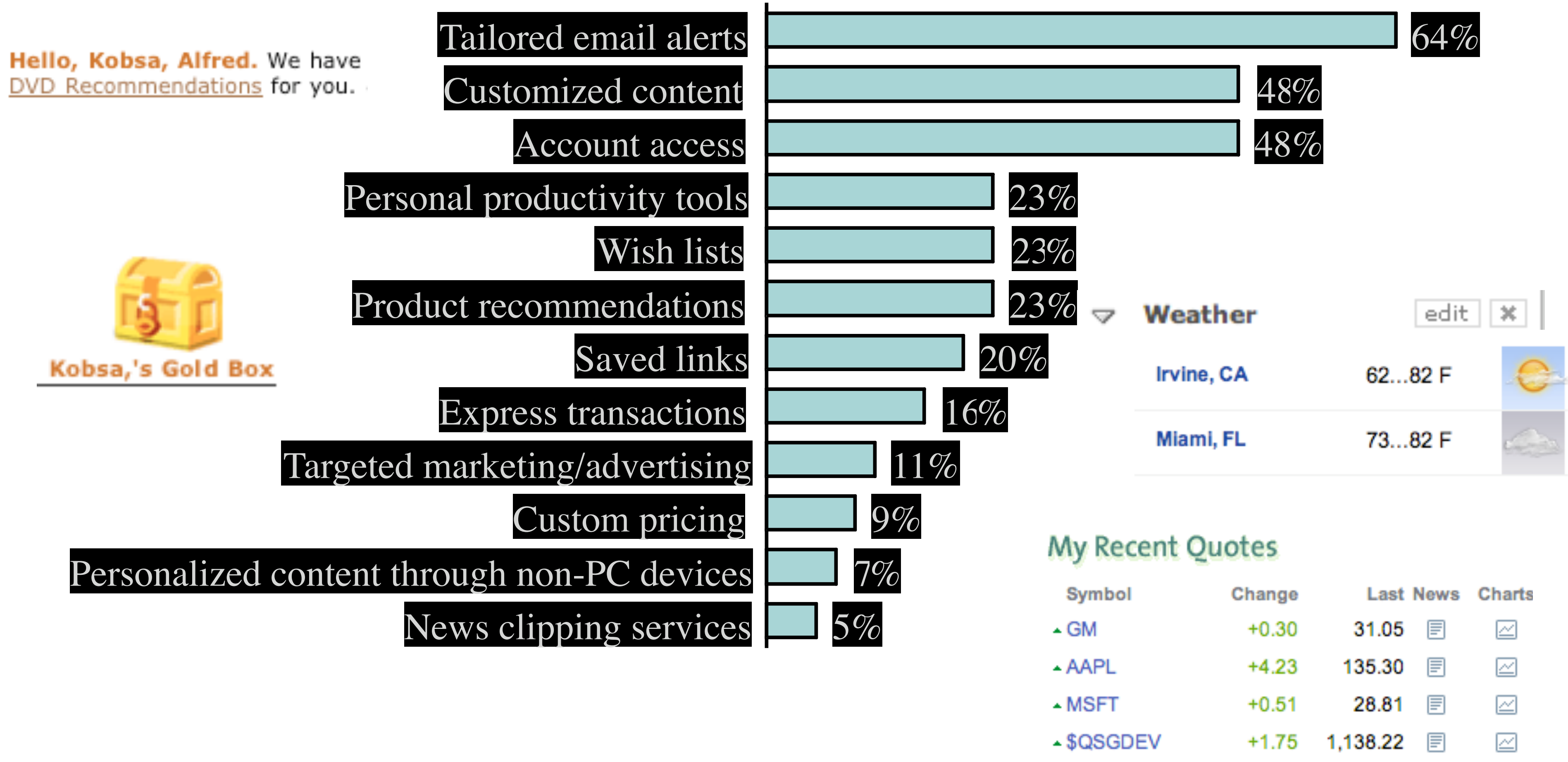
Tailoring Privacy in Personalized Systems to User Preferences and Privacy Regulations

Alfred Kobsa

Institute for Software Research
University of California, Irvine

(Joint research with Yang Wang and Max Teltzrow)

“Traditional” personalization on the World Wide Web



Percent of 44 companies interviewed (multiple responses accepted) Source: Forrester Research

Recent deployments of personalization

- Personalized search
- Web courses that tailor their teaching strategy to each individual student
- Information and recommendations by portable devices that consider users' location and habits
- Personalized news (on mobile devices)
- Product descriptions whose complexity is geared towards the presumed level of user expertise
- Tailored presentations that take into account the user's preferences regarding product presentation and media types (text, graphics, video)

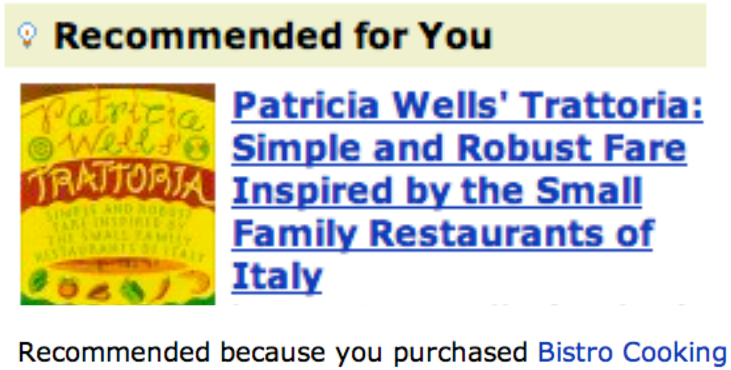
Current personalization methods (in 60 seconds)

Data sources

- Explicit user input
- User interaction logs

Methods

- Assignment to user groups
- Rule-based inferences
- Machine learning



Storage of data about users

- *Persistent* user profile
- Updated over time

Web personalization delivers benefits for both users and web vendors

Jupiter Communications, 1998: Personalization at 25 consumer e-commerce sites increased the number of new customers by 47% in the first year, and revenues by 52%.

Nielsen NetRatings, 1999:

- Registered visitors to portal sites spend over 3 times longer at their home portal than other users, and view 3 to 4 times more pages at their portal
- E-commerce sites offering personalized services convert significantly more visitors into buyers than those that don't.

Choicestream 2004, 2005:

- 80% interested in personalized content
- 60% willing to spend a least 2 minutes answering questions about themselves

***Downside**:*

Personalized sites collect significantly more personal data than regular websites, and do this often in a very inconspicuous manner.

Many computer users are concerned about their privacy online

Number of users who reported:

- *being extremely or very concerned about* ***divulging*** *personal information online:*
  67% (Forrester 1999), 74% (AARP 2000)
- *being (extremely) concerned about* ***being tracked*** *online:*
  77% (AARP 2000)
- ***leaving*** *web sites that require registration information:*
  41% (Boston Consulting 1997)
- *having entered* ***fake*** *registration information:*
  40% (GVU 1998), 27% (Boston Consulting 1997), 32% (Forrester 1999)
- *having* ***refrained from shopping*** *online due to privacy concerns, or* ***bought less****:*
  32% (Forrester 1999), 32%  35%  54%  : IBM 1999, 24% (AARP 2000)
- *wanting internet sites* ***ask for permission*** *to use personal data: 81% (Pew 2000)*
- *being willing to give out personal data for getting something* ***valuable in return****:*
  31% (GUV 1998), 30% (Forrester 99), 51% (Personalization Consortium)

Privacy surveys do not predict people's privacy-related actions very well

Surveys generally, and privacy surveys in particular, suffer from the "talk is cheap" problem. It costs a consumer nothing to express a desire for a law to protect privacy.

After all, who would not state that he is "concerned" in some sense about privacy?

Harper and Singleton, 2001
Personalization Consortium

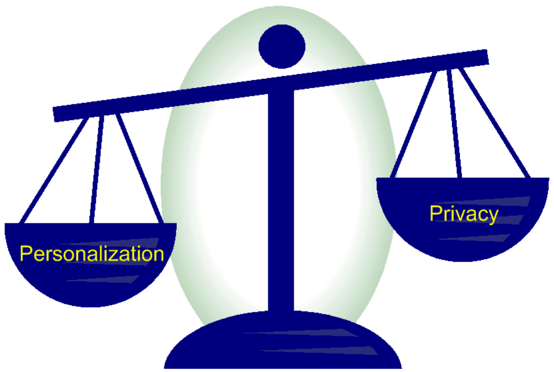
- In several privacy studies in E-commerce contexts, discrepancies have already been observed between users stating high privacy concerns but subsequently disclosing personal data carelessly.
- Several authors therefore challenge the genuineness of such reported privacy attitudes and emphasize the need for *experiments* that allow for an observation of actual online disclosure behavior.

Either Personalization or Privacy?

Balancing Privacy with Personalization

Personal data of computer users are indispensable for personalized interaction

Computer users are reluctant to give out personal data



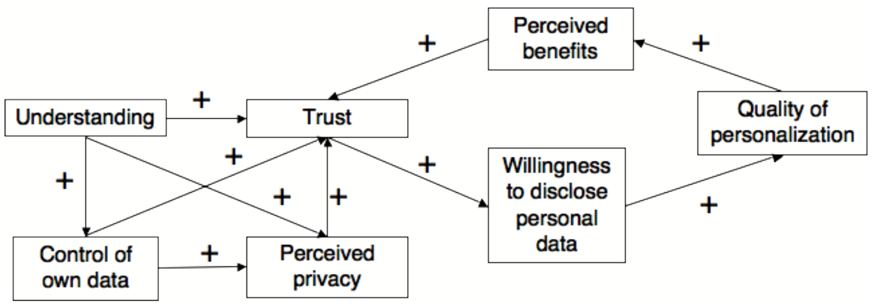
☛ **Tradeoff between privacy and personalization?**

print this article | e-mail a colleague

Personalization Vs. Privacy Debate Heating Up

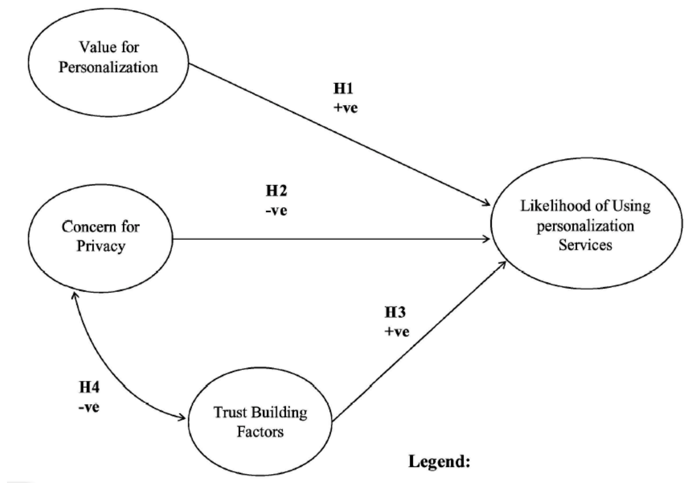
› › › ClickZ News

The tension between privacy and personalization is more complex than that…



- Indirect relationship between privacy and personalization
- Situation-dependent
- Many mitigating factors

People use complex "privacy calculus" to decide whether or not to disclose personal data, e.g. for personalization purposes



Privacy-Enhanced Personalization

Can we have good personalization and good privacy at the same time?

How can personalized systems maximize their personalization benefits, while at the same time being compliant with the privacy constraints that are in effect?

Privacy constraints, and how to deal with them

Privacy constraints

A. People's privacy preferences in a given situation
(and factors that influence them)

B. Privacy norms (laws, self-regulation, principles)

Reconciliation of privacy and personalization

1. Use of privacy-enhancing technology
2. Privacy-minded user interaction design

Privacy norms

- Privacy laws
  More than 40 countries worldwide
- Industry self-regulations
  Companies, industry sectors (NAI)
- Privacy principles
  - supra-national (OECD, APEC)
  - national (Australia, Canada, New Zealand…)
  - member organizations (ACM)

☛ Several privacy norms disallow a number of frequently used personalization methods (unless the user’s consents to them)

Privacy laws and regulations restrict the permissibility of personalization methods

- Usage logs must be deleted after each session
- Usage logs of different services may not be combined
  (except for accounting purposes)
- User profiles are permissible only if pseudonyms are used.
  (Profiles retrievable under pseudonyms shall not be combined with data relating to the bearer of the pseudonym.)
- No fully automated individual decisions are allowed that produce legal effects concerning the data subject or significantly affect him and which are based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.
- Anonymous or pseudonymous access and payment must be offered if technically possible and reasonable.
- Users must be able to withdraw their consent on processing traffic or location data at any time

Existing approaches for catering to privacy constraints

- Largest permissible dominator (e.g., Disney)
  - Infeasible if a large number of jurisdictions are involved, since the largest permissible denominator would be very small
  - Individual preferences not taken into account
- Different country/region versions (e.g., IBM)
  - Infeasible as soon as the number of countries/regions, and hence the number of different versions of the personalized system, increases
  - Individual preferences not taken into account
- Anonymous personalization (users are not identified)
  - Nearly full personalization possible
  - Harbors the risk of misuse
  - Slightly difficult to implement if physical shipments are involved
  - Practical extent of protection unclear
  - Individual user preferences not taken into account

User modeling methods

user modeling component	methods used	data used		
		demographic data	user-supplied data	visited pages
UMC_1	clustering	X		
UMC_2	rule-based reasoning		X	
UMC_3	fuzzy reasoning with uncertainty		X	
UMC_4	rule-based reasoning	X	X	
UMC_5	fuzzy reasoning with uncertainty	X	X	
UMC_6	incremental machine learning		X	X
UMC_7	one-time machine learning across sessions		X	X
UMC_8	one-time machine learning + fuzzy reasoning with uncertainty	X		X

Different methods differ in their data requirements, quality of predictions, and also their *privacy implications*

Our approach

Develop a mechanism that dynamically selects those user modeling methods that *comply with the currently prevailing privacy constraints:*

- the user's individual privacy preferences
- the privacy norms that apply to the user

User modeling methods

user modeling component	methods used	data used		
		demographic data	user-supplied data	visited pages
UMC_1	clustering	X		
UMC_2	rule-based reasoning		X	
UMC_3	fuzzy reasoning with uncertainty		X	
UMC_4	rule-based reasoning	X	X	
UMC_5	fuzzy reasoning with uncertainty	X	X	
UMC_6	incremental machine learning		X	X
UMC_7	one-time machine learning across sessions		X	X
UMC_8	one-time machine learning + fuzzy reasoning with uncertainty	X		X

Different methods differ in their data requirements, quality of predictions, and also their *privacy implications*

Product line architecture

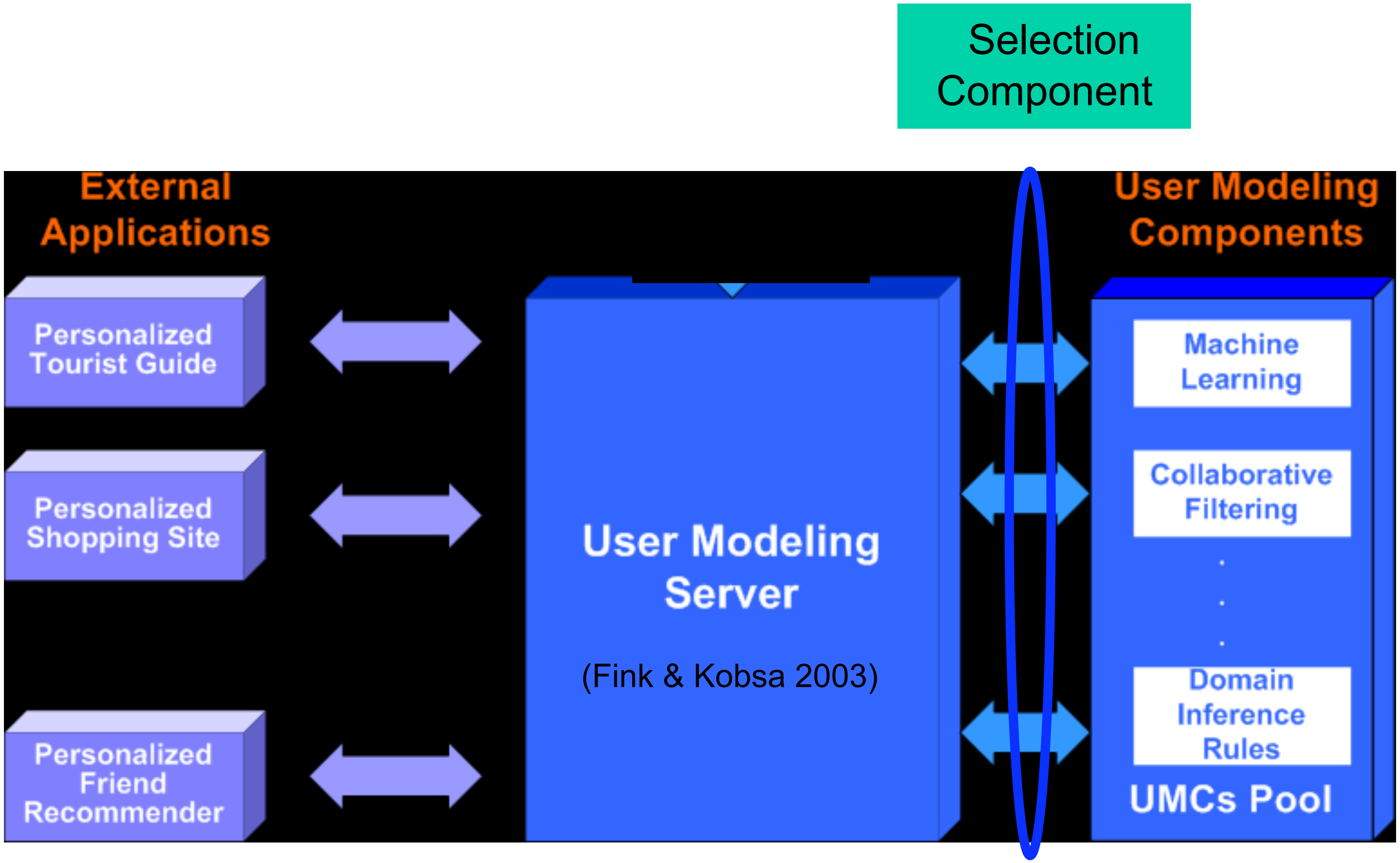
"The common architecture for a set of related products or systems developed by an organization." *[Bosch, 2000]*

A PLA includes

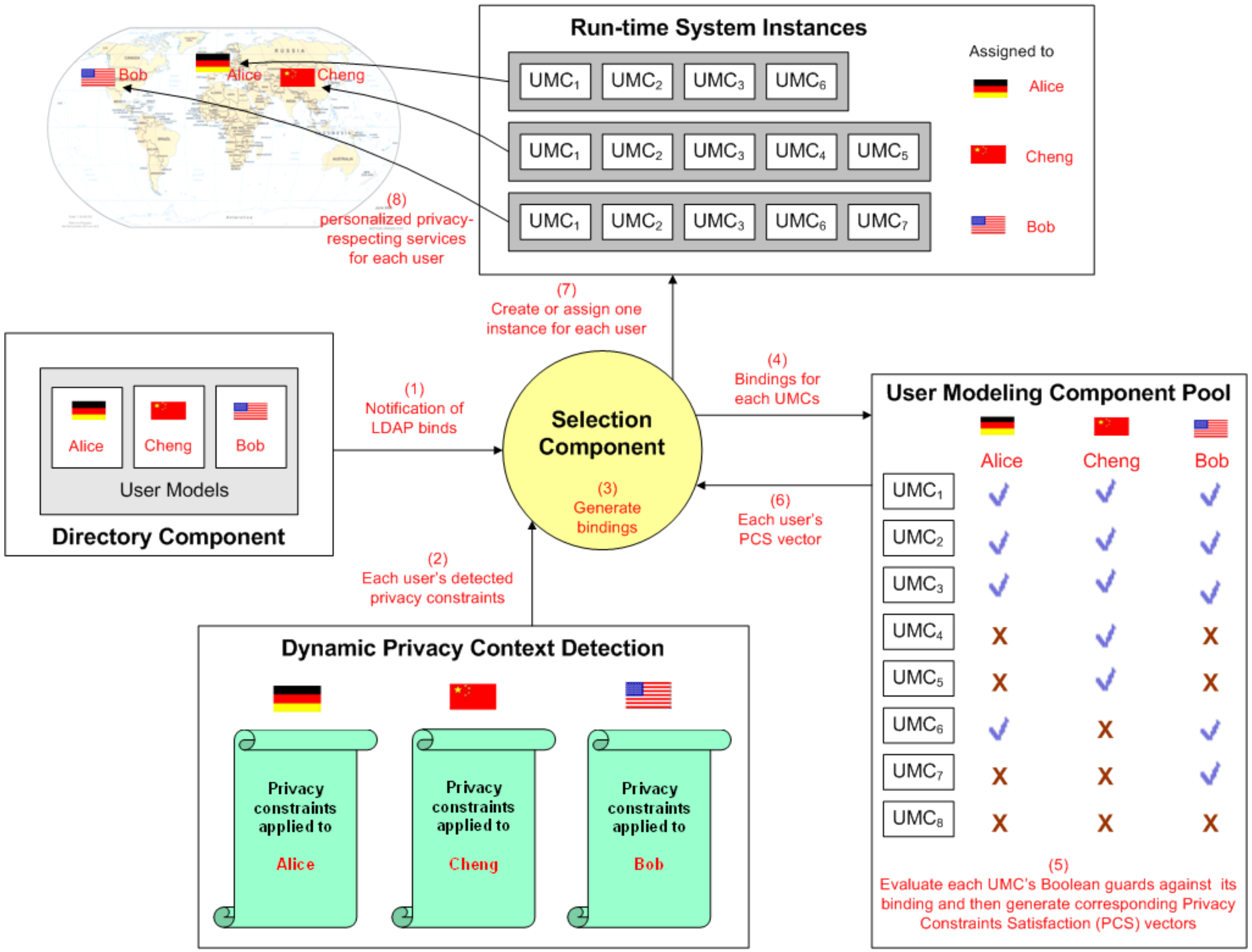
- Stable core: basic functionalities
- Options: optional features/qualities
- Variants: alternative features/qualities

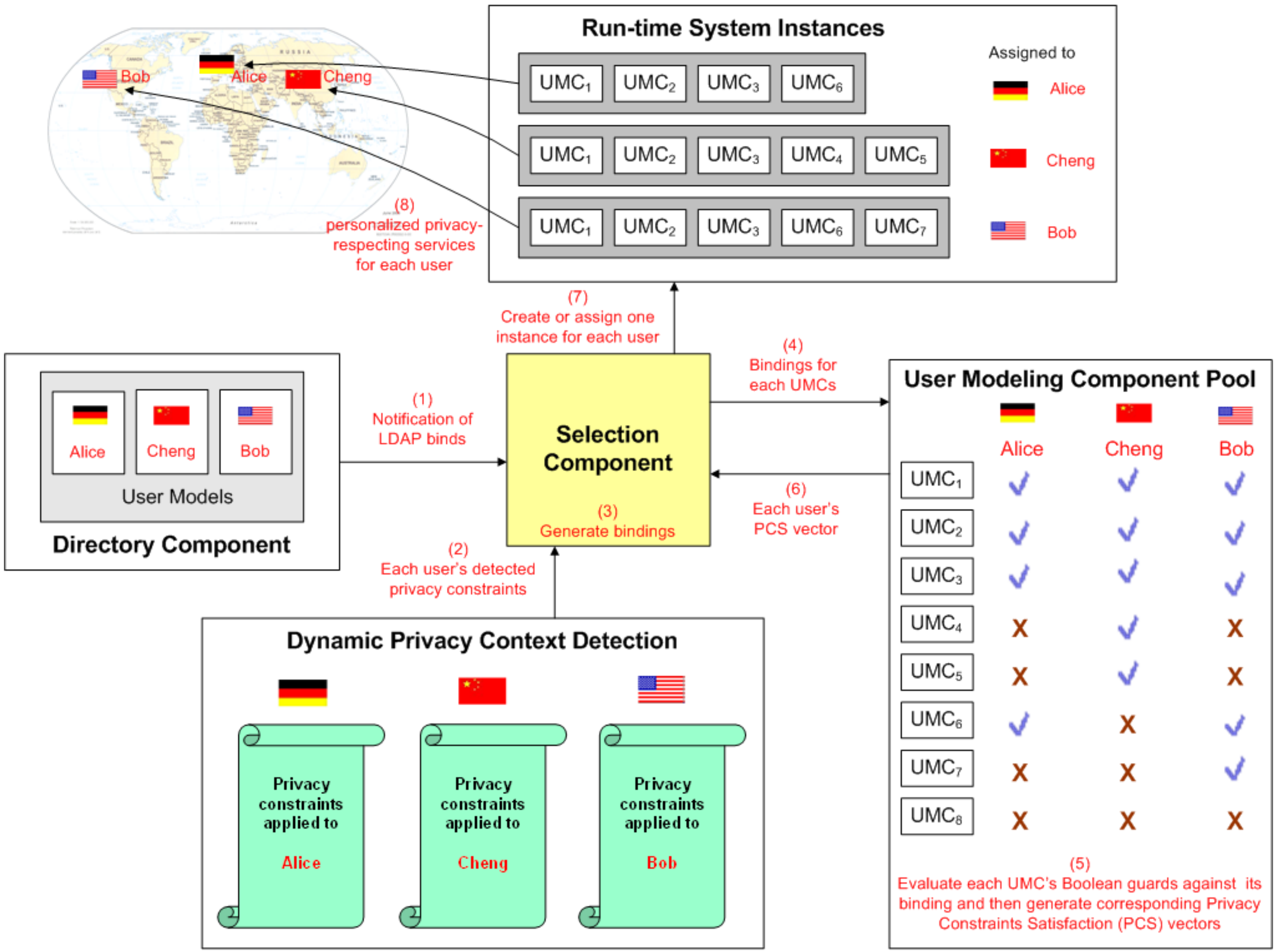
Dynamic runtime selection (van der Hoek 2002):
A particular architecture *instance* is selected from the product-line architecture

Our approach



Example: ogle.com *cum* privacy





Run-time System Instances
UMC1 UMC2 UMC3 UMC6
UMC1 UMC2 UMC3 UMC4 UMC5
UMC1 UMC2 UMC3 UMC6 UMC7
Assigned to
Alice
Cheng
Bob
Bob
Alice
Cheng
(8)
personalized privacy-respecting services for each user
(7)
Create or assign one instance for each user
(4)
Bindings for each UMCs
User Modeling Component Pool
Alice
Cheng
Bob
UMC1
UMC2
UMC3
UMC4
UMC5
UMC6
UMC7
UMC8
X
(5)
Evaluate each UMC's Boolean guards against its binding and then generate corresponding Privacy Constraints Satisfaction (PCS) vectors
Alice
Cheng
Bob
User Models
Directory Component
(1)
Notification of LDAP binds
Selection Component
(3)
Generate bindings
(6)
Each user's PCS vector
(2)
Each user's detected privacy constraints
Dynamic Privacy Context Detection
Privacy constraints applied to
Alice
Privacy constraints applied to
Cheng
Privacy constraints applied to
Bob

The privacy constraints

Privacy constraints applied to Alice

German Tele-Service Data Protection Law

Section 4(2)-4(4): profiling
Combining user profiles retrievable under pseudonyms with data relating to the bearer of the pseudonym, is prohibited.

Personal data to be erased immediately after each session except for very limited purposes.

.
.
.

Privacy constraints applied to Cheng

Cheng's own privacy preferences:

"Dislike being tracked"

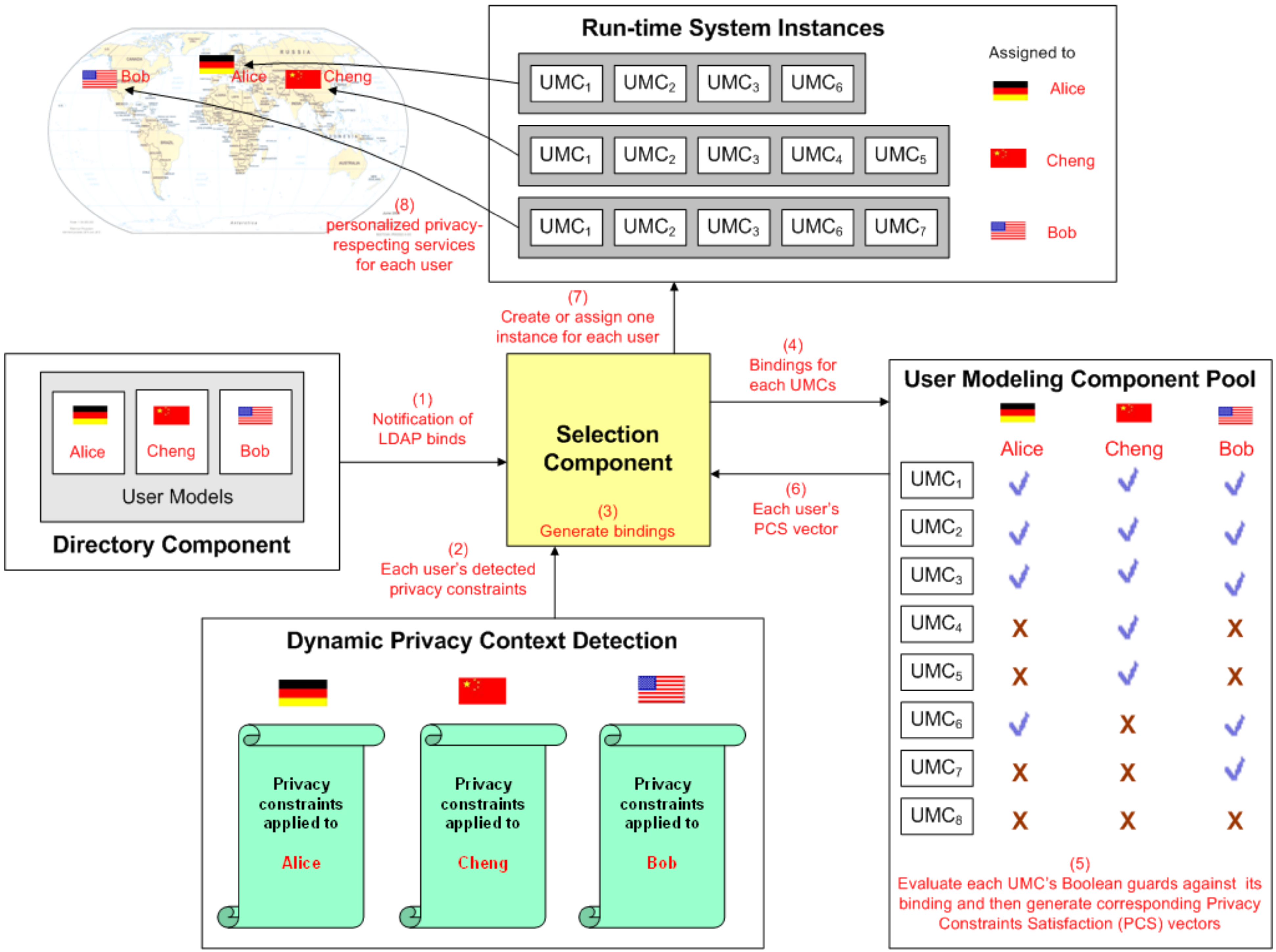
.
.
.

Privacy constraints applied to Bob

Network Advertising Initiative (NAI) Self-Regulatory Principals

Section II: NAI's Statement of Purposes
Merging non-personally identifiable use data with personally identifiable demographic data, is prohibited unless user give prior affirmative consent.

.
.
.



Run-time System Instances
Assigned to
Alice
Cheng
Bob
UMC1 UMC2 UMC3 UMC6
UMC1 UMC2 UMC3 UMC4 UMC5
UMC1 UMC2 UMC3 UMC6 UMC7
Bob
Alice
Cheng
(8)
personalized privacy-respecting services for each user
(7)
Create or assign one instance for each user
(4)
Bindings for each UMCs
User Modeling Component Pool
Alice
Cheng
Bob
UMC1
UMC2
UMC3
UMC4
UMC5
UMC6
UMC7
UMC8
X
(5)
Evaluate each UMC's Boolean guards against its binding and then generate corresponding Privacy Constraints Satisfaction (PCS) vectors
Alice
Cheng
Bob
User Models
Directory Component
(1)
Notification of LDAP binds
Selection Component
(3)
Generate bindings
(6)
Each user's PCS vector
(2)
Each user's detected privacy constraints
Dynamic Privacy Context Detection
Privacy constraints applied to Alice
Privacy constraints applied to Cheng
Privacy constraints applied to Bob

There is no magic bullet for reconciling personalization with privacy

Effort is comparable to

… making systems secure

… making systems fast

… making systems reliable

Privacy-Enhanced Personalization: need for a process approach

1. Gain the user's trust
   - Respect the user's privacy attitude (and let the user know)
     - Respect privacy laws / industry privacy agreements
   - Provide benefits (including optimal personalization within the given privacy constraints)
   - Increase the user's understanding (don't do magic)
   - Use trust-enhancing methods
   - Give users control
   - Use privacy-enhancing technology (and let the user know)

2. *Then be patient, and most users will incrementally come forward with personal data / permissions if the usage purpose for the data and the ensuing benefits are clear and valuable enough to them.*

Roadmap for Privacy-Enhanced Personalization Research

- Study the impacts of privacy laws, industry regulations and individual privacy preferences on the admissibility of personalization methods
- Provide optimal personalization while respecting privacy constraints
- Apply state-of-the-art industry practice for managing the combinatorial complexity of privacy constraints

Readings…

- A. Kobsa: Privacy-Enhanced Web Personalization. In P. Brusilovsky, A. Kobsa, W. Nejdl, eds.: The Adaptive Web: Methods and Strategies of Web Personalization. Springer Verlag.

- A. Kobsa: Privacy-Enhanced Personalization. *Communications of the ACM*, Aug. 2007

survey of privacy laws

Privacy law - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.ics.uci.edu/~kobsa/privacy/intlprivlawsurvey.html

remember this my del.icio.us

	Registration duties	Record-keeping duties	Reporting duties	Disclosure duties at website	Duty to respect user requests for		Duty to respect veto ("o out")
Argentina	• yes	?	?	• yes	• Inspection • Correction • exception	Argentina	?
Australia	?	• yes	• yes	• yes • yes	• Inspection • Inspection	Australia	?
Austria	• yes • content	?	?	• yes	• Inspection • Rectification, Erasure	Austria	• yes
	Registration duties	Record-keeping duties	Reporting duties	Disclosure duties at websites	Duty to respect user requests for		Duty to respect veto ("o out")
Canada	?	?	?	• yes	• Inspection	Canada	?

Done