Automatic Privacy Policy Clustering

... applicable privacy preferences settings to formalise the data disclosure decisions and for visualization

IFIP
Summer School on Identity Management
Karlstad, Sweden August, 6th-10th 2007

Mike.Bergmann@tu-dresden.de
Simone.Fischer-Huebner@kau.se
Andreas Pfitzmann (pfitza@inf.tu-dresden.de)
Marit.Hansen@datenschutzzentrum.de
John_Soren.Pettersson@kau.se

Automatic Privacy Policy Clustering

- Digital life becomes reality,
  - More and more online services
  - More and more personal data is released to use these services
- Data release conditions are not transparent enough
- Web 2.0 increases the need towards effective IdM

➔but how to create the policies

Automatic Privacy Policy Clustering

- Analysis of existing application scenarios
- Definition of the necessary “Sets of Data”
- Find the common structure (Similarities/Differences)
  - Analyse of the application scenarios
  - Define the main settings
- Discussion: Scenario III as the “MAX” ?!
  - Split existing business processes into subtasks
- Example implementation

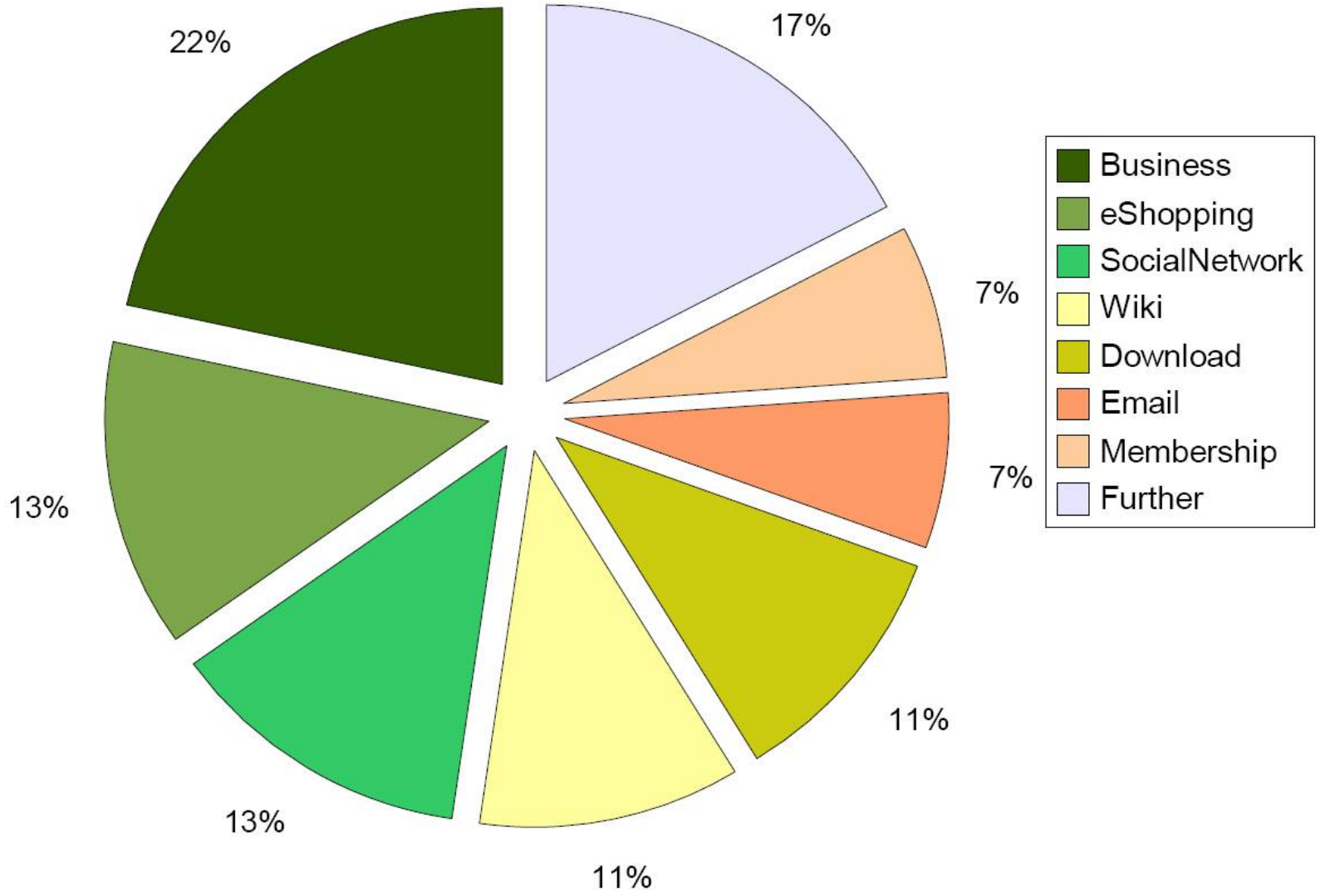
Typical Application Scenarios

- Business – prof. surrounding, full, authentic PII
- eShopping – semi-prof. surrounding, full, authentic PII
- SocialNetwork – non-prof.; no PII necessary, but released
- Download – non-prof.; no PII necessary
- Blog – non-prof.; no PII necessary, but collection becomes PII
- eMail – non-prof.; no PII necessary, but collection becomes PII
- Membership – semi-prof. surrounding, full, authentic PII

…

- Further – all others, like licensing, collaboration, news reading...

Application Scenarios - Distribution



Similarities & Differences

Scenario	*Personal Data*	*Purpose*	*Transfer*
Download	no data, not linkable		
Blog	no data, linkable		
Email	no data, linkable		
Membership	real data, linkable	current	not allowed
Business	real data, linkable	current	not allowed
eShopping	real data, linkable	current	not allowed
Social Network	(real) data, linkable	further	conditionally allowed

Derived Privacy Preferences I

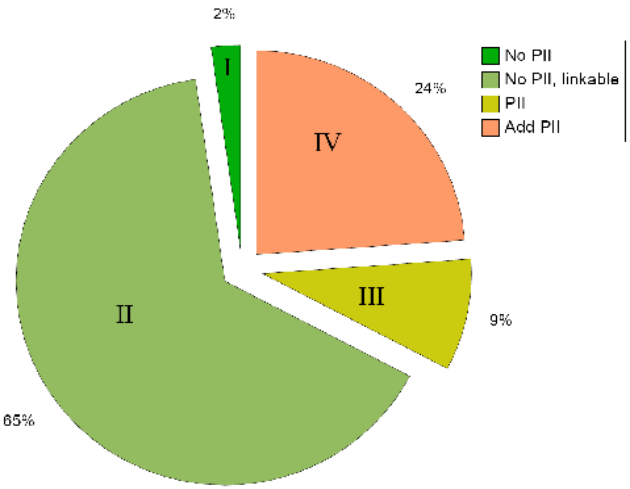
- No PII
  - Transaction pseudonyms are used, possibly linkable
  - Personal data are not released
  - Examples: weblog; create an anonymous Wikipedia entry
- No PII, but linkable
  - Use of (role–) relationship pseudonyms (not identifying the user)
  - Examples are web mailers, news panels
  - Difficult/impossible for the user to keep PII secret over time

Derived Privacy Preferences II

- Disclose necessary PII
  - Minimal amount of PII (not sensitive) binded to dedicated purpose
  - Strict no further transfer policy
  - Data release only to "trusted" partners
  - Explicit user consent
  - Example is to book a book online
- Disclose additional PII (related to III)
  - Add. (not sensitive) PII for add Services beside the primary service.
  - Data release only to "trusted"partners
  - Explicit user consent
  - Transfer to "trusted" recipients only
  - Example: customer care program

Summary

	PII	Relationship	Purpose and Transfer
I	no PII	anonymous	none (not important)
II	no PII, but user name, password, further additional non–identifying personal data	pseudonymous	only for current purpose and no transfer
III	no sensitive PII	pseudonymously or real identity	only for current purpose and strict no further transfer
IV	additional (but not sensitive) PII	pseudonymously or real identity	for additional purposes and strict no further transfer



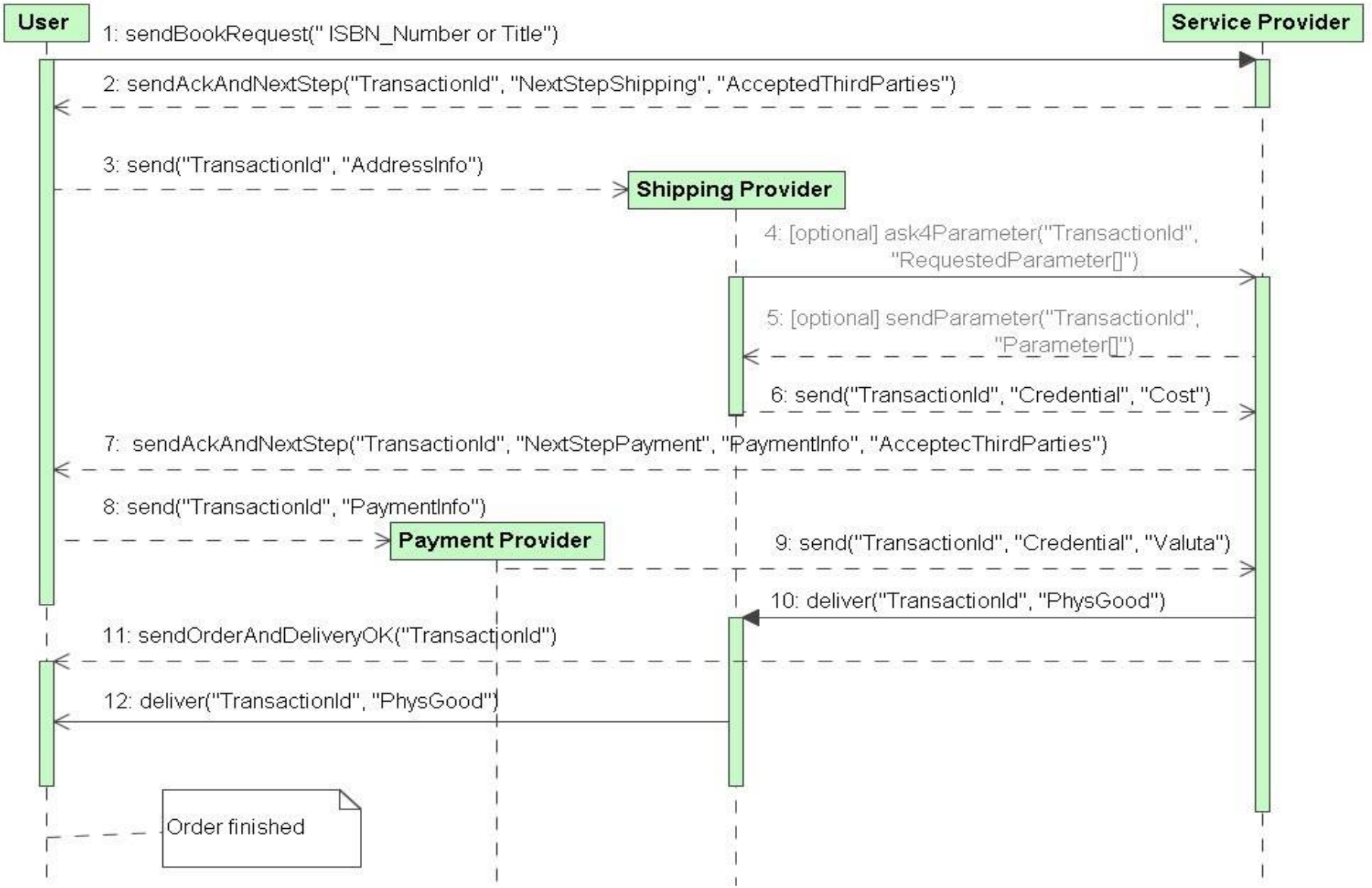
Discussion - Scenario III as the “MAX” ?!

- Transfer:
  - Each new recipient could be seen as the one and only partner
- Purpose:
  - Each new (additional) purpose could be seen as a new service and becomes „primary“ from there
- Cluster the business process accordingly

Clustering I

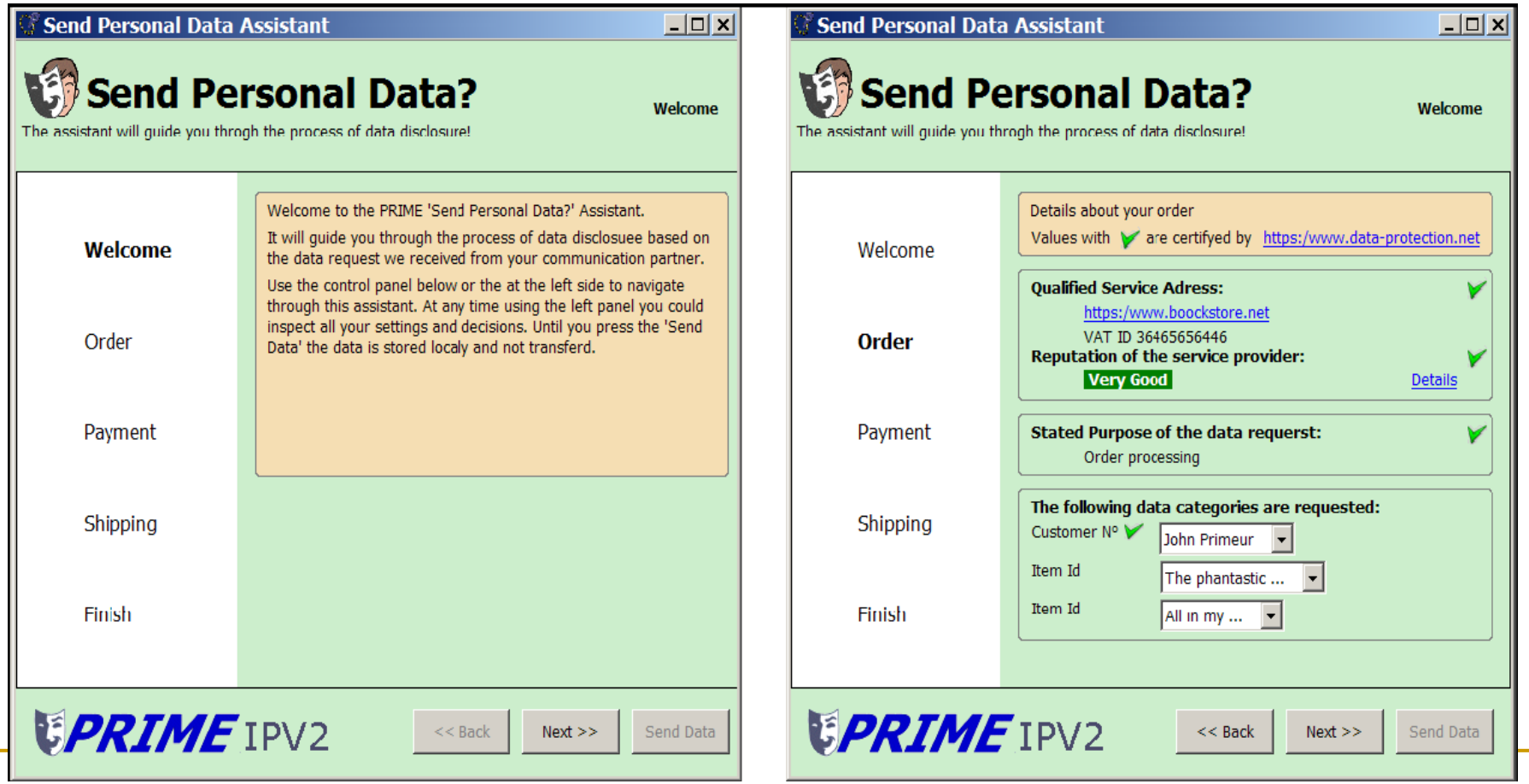
- Example for IV – Buying a Book
  - Order
  - Payment
  - Delivery
- Split it into Subtasks to achieve Scenario III
  - Order (Customer N°, ISBN; Merchant, strict no further transfer)
  - Payment (CC data, bank, strict no further transfer)
  - Delivery (Address, UPS, strict no further transfer)

Clustering II



Implementation Proposal

- Wizard like approach:



Outlook

- Find a formal description
  - „Template“ and „Preset“ as formal vehicle:
    - Template: „is a formal description of the **requirements** a certain **service provider** has to grant access to a specific **protected resource** promising an attached **data handling policy**.”
    - Preset: „ is a set of **personal data** for a dedicated template and the related **privacy preferences** for one or more specific service requests.”
- Formal protocol development to unify the clustered disclosure process
- User acceptance testing

Thanks for your attention

Send comments to mike.bergmann@tu-dresden.de