Karlsruhe Institute of Technology

# To whom flew all my personal data? Enabling users to monitor the proliferation of shared personally identifiable information

**Sebastian Labitzke**
sebastian.labitzke@kit.edu

Workshop VIII: Privacy and Transparency in the Age of Cloud Computing

STEINBUCH CENTRE FOR COMPUTING - SCC



# Outline

- Introductory part
  - Motivation
  - Related work
  - Privacy requirements in online social networks
- Contrasting enterprise and user-based identity federations
- Information Management Assistance System (IMAS)

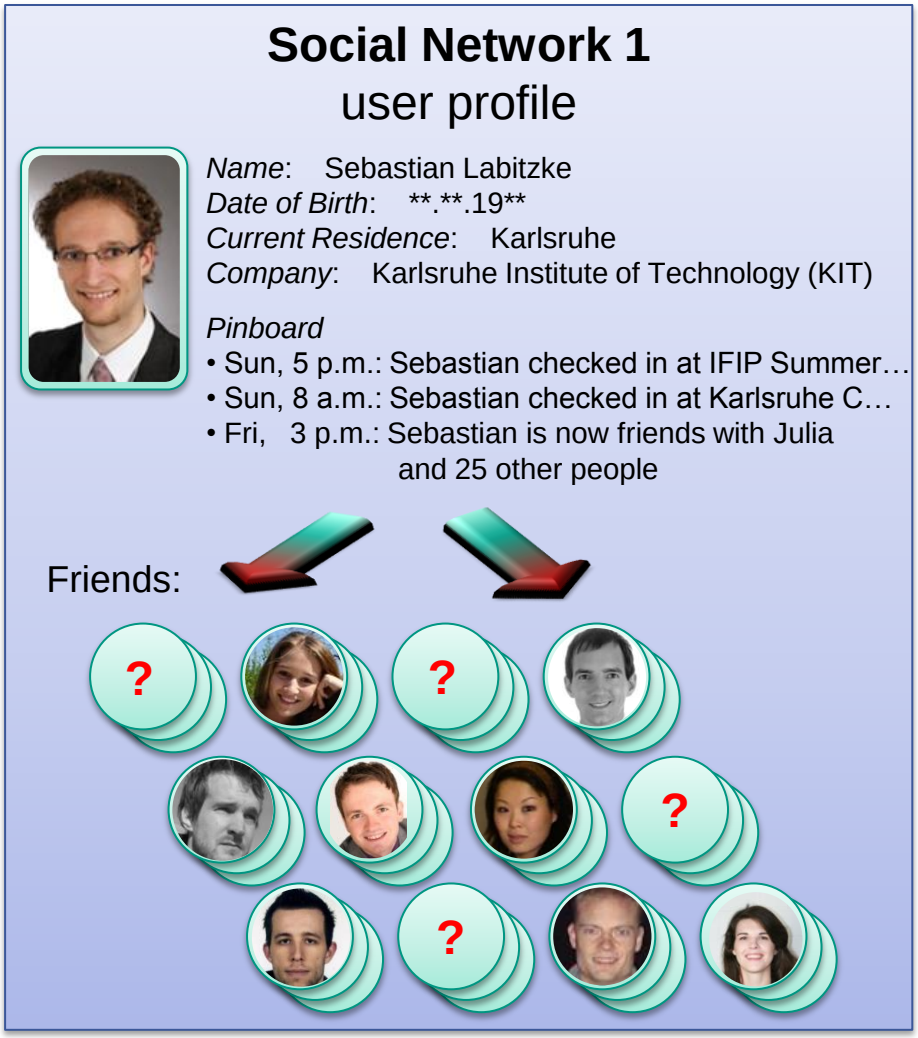
# Motivation

- Users are often careless in sharing information via online social networks (OSNs)
- Participation **versus** risk awareness w.r.t. privacy
- Risks are more and more perceived
  - E.g. due to rising media attention
  - Krishnamurthy/Wills publications, JIM studies, and others (use of privacy settings)
    - *2005: 0.06%*
    - *2007: 19%*
    - *2008: 25%*
    - *2009: <50%*
    - *2010: ~66%*

***Does this mean that these users are sufficiently aware of the risks?***

  - Restricted information is only accessible by users' friends
    - However, average number of friends in FB: 140 (standard deviation: 216)
    - Up to more than 5000 friends

# Proliferation of shared information




- Dunbar's number (**150)**
- Thoughtlessly shared information will be accessible by users who are not intended to get the shared information
- **Need of features to choose the group of receivers of shared information in a fine-grained manner**
  (cf. Google+, Clique, etc.)

→ Proactive components to control the flow of data

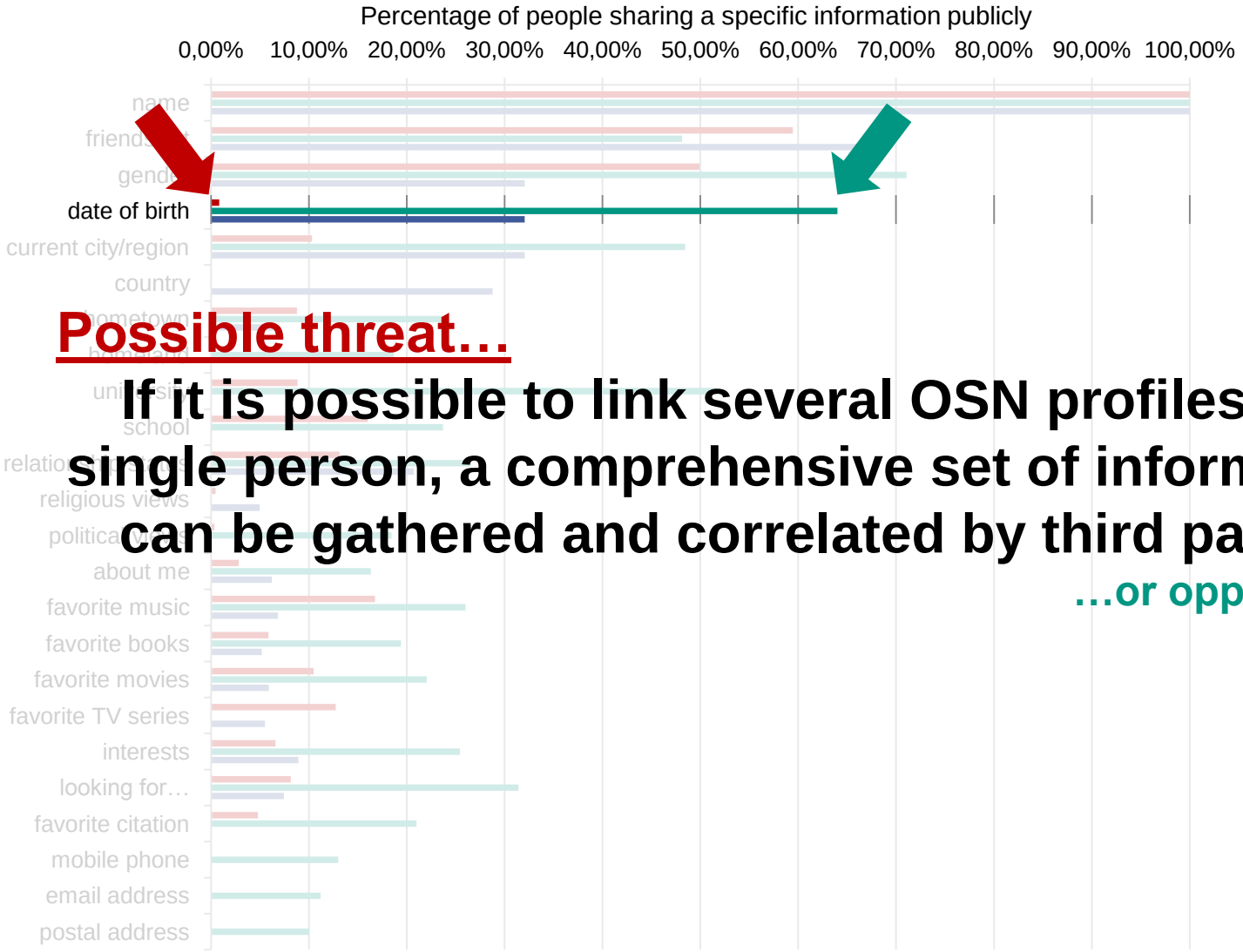
# Only proactive components needed?

- Users restrict access, but privacy settings are often not carefully used
  - Completely restricted profiles in 2011*
    - FB: 22%, SVZ: 7%, MS: 20%
  - Users share 90 pieces of content in average/month

- Concomitant threats for social network users
  - Extraction of personally identifiable information by third parties
  - Abused face recognition to identify people in images
  - Spamming, Stalking, Spear phishing
  - Unintentional information forwarding
  - Identity clone attacks
  - …

- Users are often member of several OSNs (personal, business,…)
  - although not necessarily active
  - …and often share diverse sets of information

# Diversity of publicly available information




**Possible threat…**

**If it is possible to link several OSN profiles of a single person, a comprehensive set of information can be gathered and correlated by third parties**

**…or opportunity?**

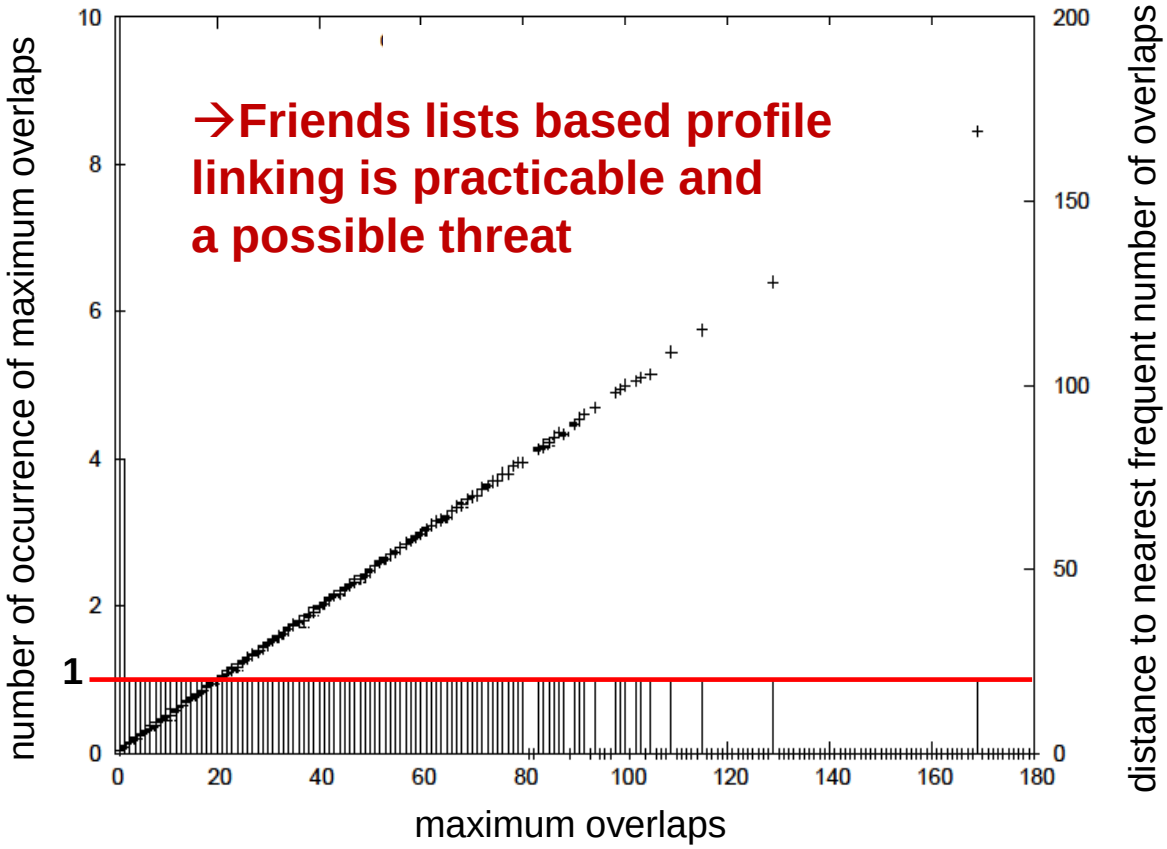
# Profile Linking

- “Linkability” of several profiles of a single user
  - Third parties are able to gather lots of rather distributed information
  - With a **linkability investigation*** we demonstrated that…
    - Profile linking is practicable on the basis of publicly available information
    - Profiles can be linked even at low cost
    - Friends lists based profile linking is sufficient to identify and correlate several social network profiles of a single natural person

* **What Your Friends Tell Others About You: Low Cost Linkability of Social Network Profiles**
S. Labitzke, I. Taranu, H. Hartenstein, In Proceedings of the 5th International ACM Workshop on Social Network Mining and Analysis SNA KDD, San Diego, CA, USA, Aug. 2011

# Analysis: comparisons of friends lists*

In each comparison a profile from network A was compared to n profiles from network B




* **What Your Friends Tell Others About You: Low Cost Linkability of Social Network Profiles**
S. Labitzke, I. Taranu, H. Hartenstein, In Proceedings of the 5th International ACM Workshop on Social Network Mining and Analysis SNA KDD, San Diego, CA, USA, Aug. 2011

# Diversity of shared information

## Social Network 1
user profile (business)

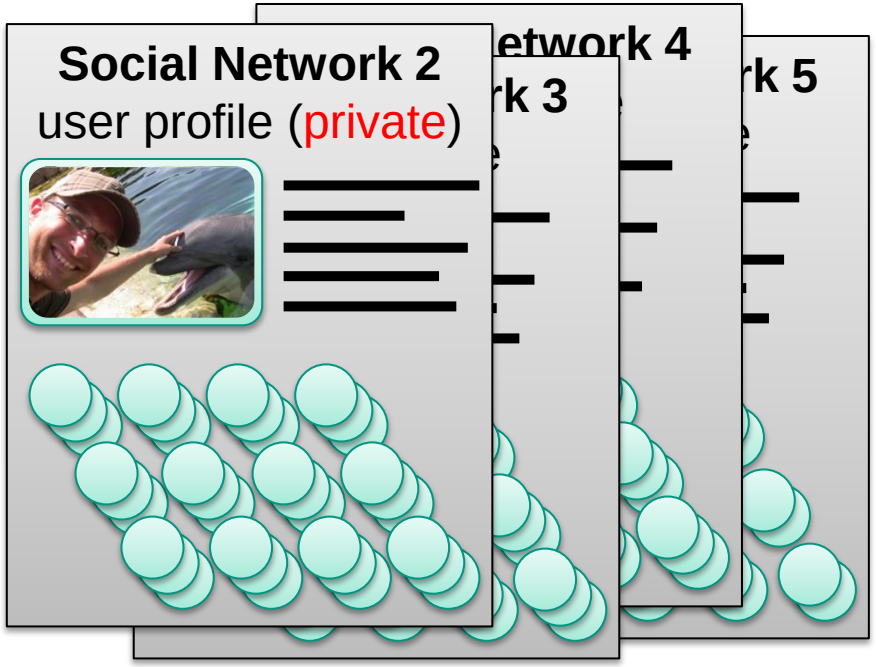


*Name*: Sebastian Labitzke
*Date of Birth*: **.**.19**
*Current Residence*: Karlsruhe
*Company*: Karlsruhe Institute of Technology (KIT)

*Pinboard*
- Sun, 5 p.m.: Sebastian checked in at IFIP Summer…
- Sun, 8 a.m.: Sebastian checked in at Karlsruhe C…
- Fri, 3 p.m.: Sebastian is now friends with Julia and 25 other people

Colleagues and Friends:






→ **Need of reactive features to monitor the data proliferation**

# Requirements at a glance

- Proactive components
  - Who will get access to which shared information

- Reactive components
  - …who has access to which personal information
  - …which information can be accessed by third parties
  - …which information can be linked to one another

  → Maintaining an overview of the availability of personal data
  → Call attention of users on the risks regarding privacy

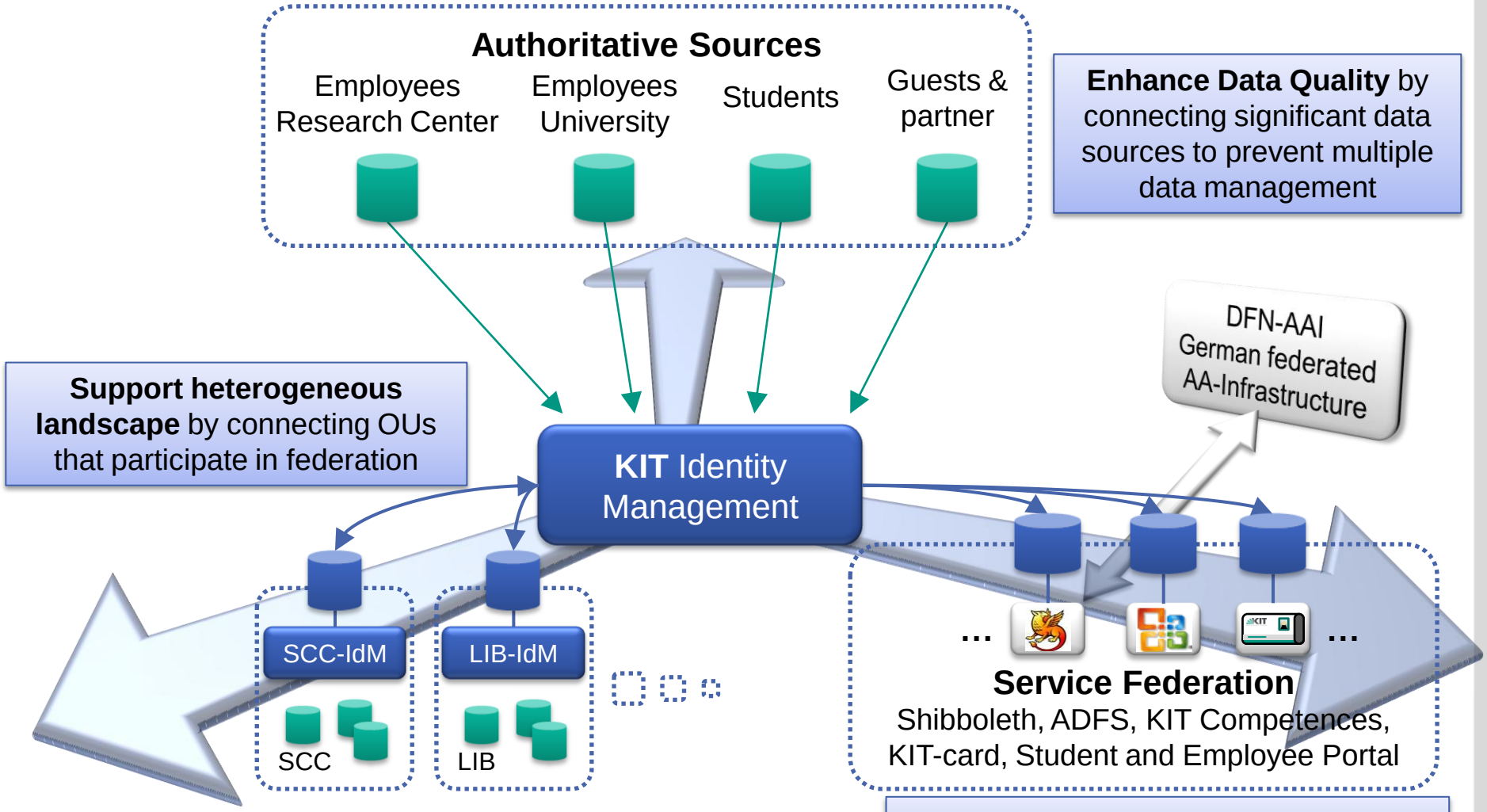
- Necessity of a system to manage (monitor and control) data flows

***Information Management Assistance System (IMAS)***

# Outline

- Introductory part
  - Motivation
  - Related work
  - Privacy requirements in online social networks
- Contrasting enterprise and user-based identity federations
- Information Management Assistance System (IMAS)

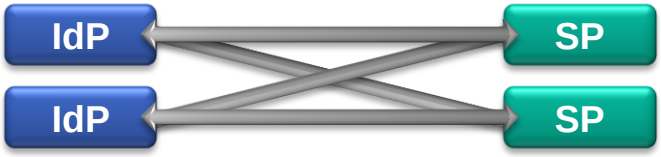
# Federated Identity Management (FIM) at KIT



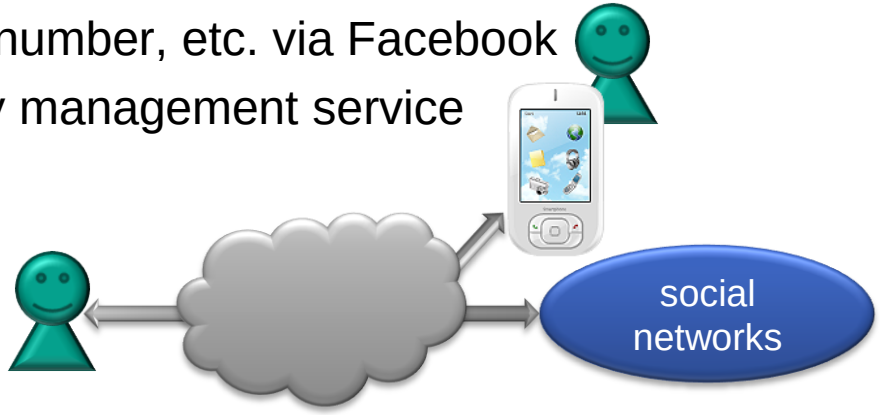
# User-based identity federations

- From communication services to self-projection opportunities
  - Federated identity services do not have to be “provider-engineered”
  - “Social services”: easier to access and ubiquitous by default
  - Users’ behavior and demands are more relaxed in social networks than in provider-engineered environments

- Example of an existing **user-based identity federation**
  - Linking of Facebook friends and contacts on smart mobile phones
  - → Updates of email address, phone number, etc. via Facebook
  - → Federated and user-based identity management service

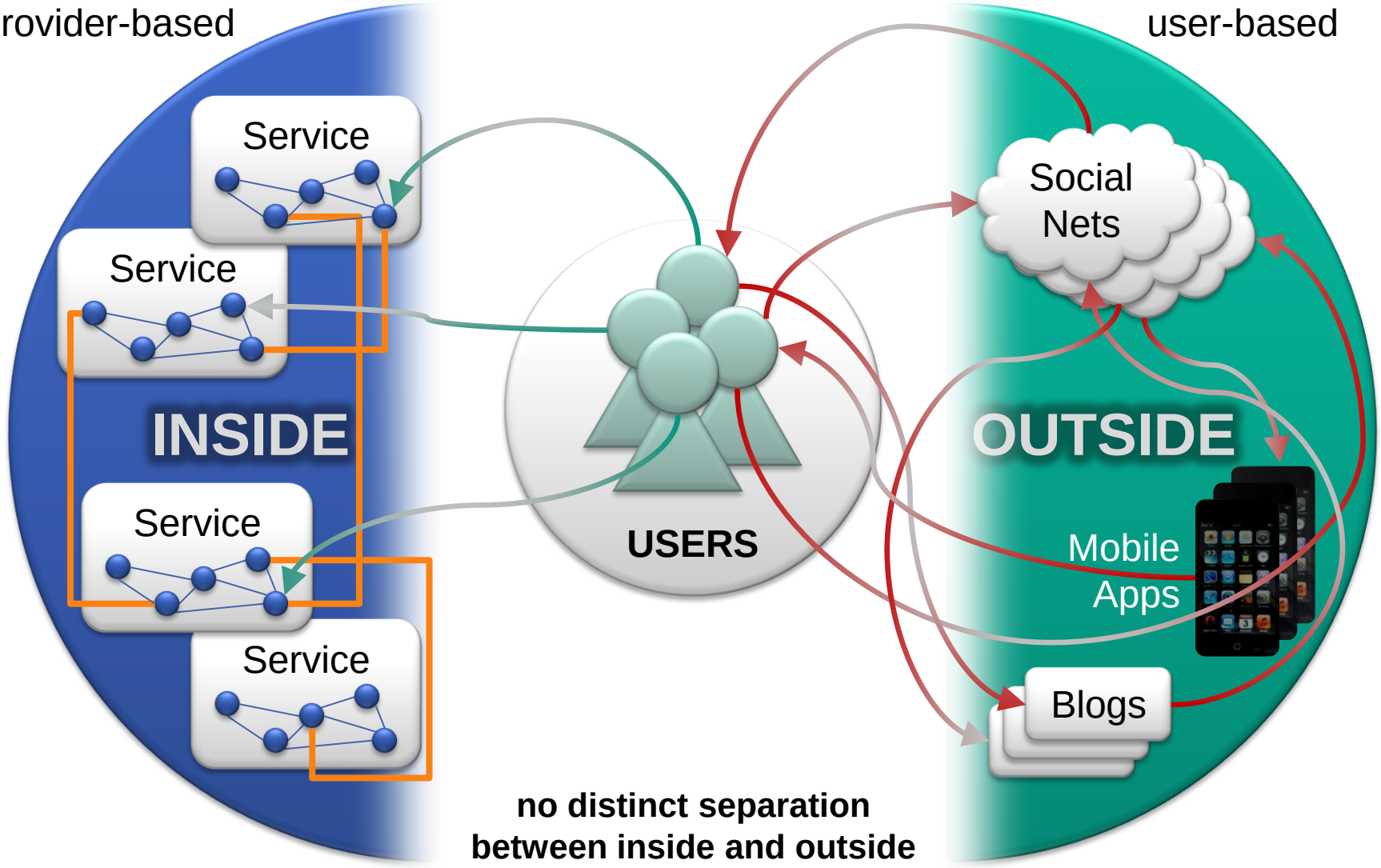



Today, most federations are provider-based




Federations can be user-based, for instance, in light of online social networks

# Enterprise vs. OSN environments



# “Inside” vs. “Outside”

→ Provider-engineered processes
- Developers implement processes
- Administrators are able to monitor data flows
- Often many rules and restrictions → takes time
  (data protection acts, communication & coordination overhead,…)

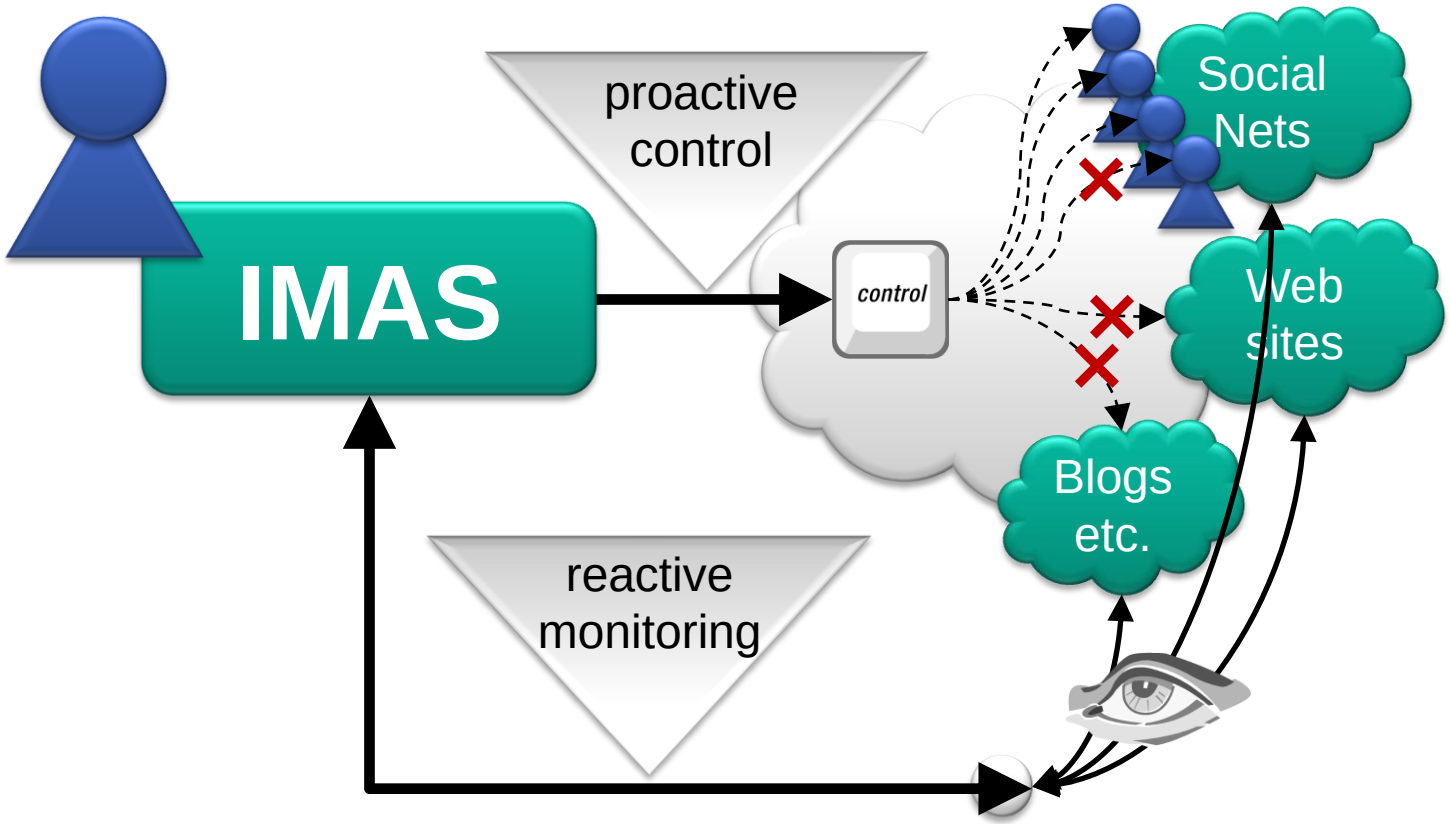
- However, “user-engineered” processes require technical support of users (such as competences of a provider and its developers and administrators)
  - Monitoring, control and management of information flows

- Access restrictions regarding receivers of a shared information are not sufficient
  - Reactive components should be provided
  - Monitoring of the availability of personal identifiable information *after* sharing
  - Revealing risks with respect to publicly available information of a specific user

# Outline

- Introductory part
  - Motivation
  - Related work
  - Privacy requirements in online social networks
- Contrasting enterprise and user-based identity federations
- Information Management Assistance System (IMAS)

# Information Management Assistance System



# IMAS requirements

- Requirements of an IMAS are very similar to requirements in the field of enterprise identity management
  - Just developer and administrator roles are not separated
- **Proactive Control**
  - Capability to decide who will get access to shared information
- **Reactive Monitoring**
  - Capability to monitor the proliferation of personal data

**Comparison to enterprise identity management systems**

Developer's capabilities
→ Support of "process design"

Administrator's view
→ Monitoring of data flows

# IMAS Building Blocks – Proactive Control

- More or less known
  → UMA, OAuth, OpenID, Information Card, etc.
  → Clique, etc.

- User are often challenged by adjusting provided settings
  → UCAID* (policy-based support of users regarding attribute releases)

- *Users are able to allow or deny that third parties will get requested information*

- Sufficient proactive components are not provided
  - Tools only show that data is to be shared
  - Implications of adjustable permissions often not shown
  - Appropriation of shared data is only, if at all, implicitly comprehensible
  - Third party providers often request the majority of potentially available information
  - Users are not able to pick and choose which piece of information will be shared
  - Users can use services if they admit access to every requested information

* **User-Controlled Automated Identity Delegation**, T. Höllrigl, H. Kühner, J. Dinger, H. Hartenstein, In Proceedings of the 6th IEEE/IFIP International Conference on Network and Service Management CNSM 2010, pp. 230-233, Niagara Falls, Canada, Oktober 2010

# IMAS Building Blocks – Reactive Monitoring

- Today, reactive components are not implemented
- Maintaining an overview of proliferating information: not possible, today

- **Benefits from profile linkability**
  - Threat: Aggregation of user's information shared in different social networks
    - Low cost linkability without complex correlation or mining algorithms
    - Comparisons of friends lists suitable
  - **Vice versa, profile linking to demonstrate users their own virtual appearance across several social networks**
    - How much information the user reveal to third parties
    - Bold and simple insights in information that could be linked to one another
    - Psychological point of view: see *

…just a building block

* **How I and others can link my various social network profiles as a basis to reveal my virtual appearance**, S. Labitzke, J. Dinger, H. Hartenstein, In Lecture Notes in Informatics (LNI - Proceedings, GI-Edition), 4. DFN Forum Kommunikationstechnologien, Bonn, Germany, Juni 2011

# Discussion on an IMAS and its requirements

- Are disadvantages associated to such a system?
  - → Possible threat: use of IMAS on information of arbitrary users
    - Would reveal the proliferation of shared information of this user
    - Would empower to link this information to one another
  - → Also questionable: how IMAS could be implemented on a compliant basis
    - Probably, by detecting personal data of a single user, shared information of other users have to be analyzed
    - Permission facts

- Where to deploy?
  - OSN Application(s)
  - Web OS

- However,
  - IMAS constitutes a facility that might help users to keep track of their information
  - IMAS could motivate to adjust privacy settings more carefully

# Summary

- In the future
  - **Provider-based AND user-based federations**

- Requirement: *EMPOWER THE END-USER*
  - Monitor and control data flows
  - Proactive *and* reactive components

➢ **Information Management Assistance System**

- *Such system should enable OSN users to control to whom their shared data will flow before they share information and to monitor to whom it flew afterwards*

*sebastian.labitzke@kit.edu as well as via*  *and*  *and… ;-)*