User Representation in eCommerce and Collaboration Applications

Michael Koch
Department of Informatics, Technische Universität München, Germany
KochM@in.tum.de

Kathrin Möslein
TUM Business School, Germany
Moeslein@ws.tum.de

Abstract

The development of the Internet was originally based on the assumption that a user remains anonymous. However, more and more services need to know the user for providing personalized services or for presenting the user to other users. As in real life, a user will interact with different services hosted by different providers. With the current approach users have to provide and update information about their identity and interests for each service independently. That results in cold-start problems for new services and inconvenience for the user. In this paper we argue that user-centric global identity management is needed for future e-commerce and collaboration applications. We present the current state of art in the area of identity management, discuss needs and possibilities for future developments, and show some results of the work we have done in this context.

1. Introduction

Personalization and community support are increasingly considered to be an important ingredient of successful (Web) applications for e-commerce and collaboration.

1.1 Personalization

Personalization techniques are used for tailoring information services to the needs of individual users. In marketing, personalization supports one-to-one marketing which should increase the customer share over a lifetime. What used to be possible in the corner shop, since the shopkeeper knew her customers personally, will be extensively possible in the electronic medium by the storage of profiles and the automatic evaluation on the basis
of predefined rules (Schubert & Koch 2002). In addition to online shops personalization becomes more and more important also in (Web-based) communication and collaboration services like community platforms.

Technically, personalization is about selecting or filtering information objects or products for an individual by using information about the individual (her customer profile). Different methods are known for performing this selection. These methods range from content based filtering with rules or vector similarities to automated collaborative filtering (see Schubert & Koch 2002 for more information). But independently of the personalization method the ability to deliver personalization is always based on the acquisition of a picture of the user. Depending on the personalization method used, there are different requirements to the representation of this picture. For content based filtering information about preferred content and relationships to content objects has to be stored. For collaborative filtering relationships to other users and ratings or comments have to be managed.

1.2 Community Support

Personalization is not the only reason for services to collect user profile information. More and more often Web-based services offer some kind of community support functionality. That means that the users are not supported independently from each other but are put into contact with each other. Users are supported in exchanging information, getting in contact, and communicating with each other. Bringing communities of people together stimulates three major potentials:

1. the building of trust,
2. the collection and effective use of (trusted) community information and
3. the economic impacts of accumulated buying power.

The economic impacts of communities for accumulating buying power have been discussed by Hagel and Armstrong (1997). They mainly focus on groups of Internet users that are drawn together around products or companies and use the extended possibilities of the online medium to cooperate and gain advantages they would not have if they were acting as isolated customers (better information, discounts). Those groups are often referred to as virtual communities of transaction (Schubert 1999). In addition to the accumulation of buying power this type of community is a source for valuable data about the products and about community members. User profiles can be harnessed by the operator of an electronic transaction platform with the consent of the users. In communities of transaction the additional information about the users is often the basis for personalization by making use of techniques such as collaborative filtering, data mining, and personalized user interfaces.

In addition to the use of user profiles for personalization, in community support platforms user profiles are also needed for presenting users to each other. In communication, which is the primary activity in communities, knowing the identity of those with whom you communicate is essential for understanding and evaluating an interaction and for building trust (Donath 1998). So the community members have to be aware about each other and have to know about each other. There is no need to have the identity linked to the real world identity of the user – but the identity in the community has to be persistent. Persistent pseudonymous identities can for example be found in online auction platforms like eBay where reputation information is stored as part of the identities (Kollock 1999).
User Representation in eCommerce and Collaboration Applications

In sum, for modern applications user representations have to be available for personalization and for presenting users to each other. In the remainder of this paper we will

- detail some issues of user representation and identities and highlight the main problems with user representation (Section 2)
- present a basic idea (“identity management”, see Section 3) and a technical solution (the so-called “IDRepository”) to some of the problems (Section 4), and
- review existing work to highlight the differences among existing approaches and our new approach (Section 5).

Finally, we will lay out a work agenda for user representation and identity management in the future (Section 6).

2. User Representation and Identities

The Webster English Dictionary describes the word identity as: “1) the condition or fact of being the same or exactly alike (sameness, oneness); 2a) the condition or fact of being a specific person or thing (individuality); b) the condition of being the same as a person or thing described or claimed” (Webster 1988).

In e-commerce applications the aspect of identity as proving to be a specific person plays an important role and already has been addressed broadly. For personalization and for collaboration support however, the aspect of identity as all information that describes a specific person is much more important. In this paper we therefore will regard identity more in the context of user profile, a set of information representing a user or clearly related to a user or role in the digital world. In the rest of the paper we therefore will use the terms user representation, user profile and identity synonymous.

![Figure 1: Personalization as Key to Ambient Computing (Schubert & Koch 2002)](image-url)
Figure 1 shows different types of information about a user that can be used for personalization. This information ranges from the name of the user, demographic attributes and the history of past purchases to dynamic attributes like the current position as used in location based personalization.

In the context of user representation in Web-based e-commerce and collaboration systems three issues are of major importance: 1) modeling user representation, 2) acquiring user representations and 3) managing access rights and awareness to user representations.

2.1 Modeling User Representations

For coding simple user profile information like addresses or payment information some standards are available. Examples are the vCARD standard (Howes et al. 1998) or the profile scheme included in W3Cs P3P specification (P3P 2000). These approaches are mainly based on hierarchically structured sets of attribute value pairs. For more complex information like interests or browsing histories currently personalization applications define proprietary codings based on the application and on the algorithms operating on the information. In addition to these proprietary codings used in live applications there is some work on user profiles emerging from Artificial Intelligence and Knowledge Management research. See (Essing 2001), (Fink & Kobsa 2000) or (Mertens & Höhl 1999) for more information on abstract modeling of user profiles and user profile servers.

2.2 Acquiring User Profile Information

There are various methods for capturing user profile information, which engage the user to different degrees. One usually distinguishes asking the customer (fill-in-profile, explicit feedback or ratings) and watching the customer (click-stream- or transaction-analysis). While the discussion of these methods is important it does not address some basic issues in user profile acquisition:

1. Users often do not trust services that collect and use profile information (and therefore try to provide no or false information).

2. Even if the user cooperates, some time and effort is needed before enough information is collected to provide appropriate recommendations. So, small sites do not have a chance to get enough information from the user (by watching).

The second issue is called “cold-start problem”. This means that users expect good recommendations from the beginning, but the system is only able to provide recommendations after having asked the user a lot of questions or after having watched the user for some time. This issue is of special importance in the field of Small and Medium-sized Enterprises (SMEs).

2.3 Managing Access Rights and Access Awareness

Finally, when user data is collected, there is the question of control. Users are only willing to agree to collecting and storing data about them or to provide data if they have and benefit from it, and if they have control and awareness of what is done with the data. Some current work tries to address this issue with privacy policies provided by services and privacy statements expressed by the users (see the W3Cs P3P project for one example (P3P 2000)). However, these works only address part of the picture. According to the two major usages of user profile information as described in the previous section,
access control has to address the issue of what platform should be provided with what information (to do personalization), and with which users should be able to see what information (for matchmaking and trust building). In addition to providing these two aspects of access control, user profile storage has to provide awareness about what information is stored and what it is used for.

3. Identity Management

In order to solve the key problems mentioned in the previous section (trust and cold-start), our concept builds on

- giving control of profile information back to users (to solve the trust issue) and
- allowing the reuse of profiles among the different personalization services (to solve the cold-start problem).

To do so we first separate

- the usage of user profiles (in the personalization services) and
- the storage of user profiles (e.g. in a central user profile server).

This setup opens the possibility of user profile reuse and provides a single location where user access control and user awareness (of who is using what information) can be implemented in a way the profile owners trust. However, this usage of user profiles by different applications also makes the modeling issue more important and raises the need for standards.

Taking a user-centric point of view to the issue leads us to no longer talking of user profiles and user profile management but of identities and identity management. An identity is the set of attributes describing (an aspect of) a person. Managing which information is available for which application is called identity management. Identity management is something we do in normal conversation everyday when we decide on what to tell one another about ourselves. In interactions with others we consider the situational context and the role we are currently acting in as well as the respective relationship with the interaction partners. This results in different sets of information being released to different interaction partners. Sometimes this leads to the situation that a person is known under different names in different contexts, e.g. by using special names, nicknames or pseudonyms suiting the occasion (Köhntopp & Bertold 2000).

Also or especially in the digital world people are using different (digital) identities. When interacting with different applications from different providers and using different identities it becomes hard to keep track of the information which service stores which information, and to keep the information in the services up to date.

An identity management system would allow people to define different identities, roles, associate personal data to it, and decide whom to give data and when to act anonymously. An identity management system would empower the user to maintain their privacy and control their digital identity.

Some projects that seem to follow similar concepts have been launched or announced in the past years: Microsoft Passport, Novell DigitalMe, the Liberty Alliance project and XNS. However, most of these solutions follow the centralization approach but do not give the user control of her profile (make the user owner of her profile). Additionally, these projects do not deal with complex user attributes that are needed to model interests or
relationships. They mainly focus on authentication and service centric exchange of simple attributes in contrast to authorization and awareness of access to complex user profile information. We will review related work in more detail in Section 5.

4. IDRepository

Since current solutions do not support user control in identity management we have built an identity management solution, which is clearly focused on user empowerment and which supports complex user profile attributes, the IDRepository (Koch & Wörndl 2001, Koch 2002a, 2002b).

![Identity Management Architecture Components](image)

**Figure 2: Identity Management Architecture Components.**

Our technical approach follows the ideas presented in the previous section, to separate user profile information from services that make use of it, and store the identity under the control of the identity owner in a place where it can be maintained by the user and be accessed by different services (with permission of the identity owner).

The core component in our architecture is a user profile repository service (IDRepository) that stores identities and offers the identity owner and authorized services interfaces to access this information (see Figure 2). The server offers the functionality to store more than one identity and to link identities to each other (defining data propagation paths).

For using the repository we have several possibilities placed between the following two extremes:

- one central identity server for storing all identities of all people
- one or even several servers per person storing different identities

We imagine that in the real world there will be a federated solution with independent identity providers – companies that operate identity servers – and no central authority (see Figure 3).

The services that read the profile information from the IDRepository should have a possibility to cache this information for some time. To implement this caching
functionality we need a means for keeping the cache up to date (and for the user to request deletion of the cached copy). After the negotiation of the basic lease this whole process can be seen as replication of the data with a master copy.

Figure 3: Identity Management Network.

Now that we have outlined the general architecture for storing and accessing profiles, there is the question of how a profile should be structured to be of general use and allow interoperability. As mentioned in Section 1 current personalization solutions mainly rely on proprietary user profiles or on simple standards defining attributes for name, payment and delivery information.

When reviewing information needs in Internet based (personalization) services the following types of information can be identified:

- basic and demographic attributes like “name” or “gender”
- information about interests: This can be represented by correlations with predefined clusters or stereotypes (e.g. in IFAY (www.ifay.com)), by explicit attributes (e.g. “interest.music = ‘hip hop’”) or by collaborative interest definitions (correlations with other users). The source for all of this can be ratings given by the user to information (implicit by visit or explicit).
- ratings given by the profile owner to information items or products
- information about relationships to other people: colleagues, buddies
- browsing and shopping (transaction) history
- preferences
- PIM (personal information manager) information (e.g. calendar)

Some of this information can be stored in a standard way using attribute value pairs with string values, but not all of it. Therefore, our approach extends the standard approach by new data types and domain ontologies for attribute names mainly for expressing interests, ratings and for relationships. Additionally, there is the possibility to have multiple values in any place in the hierarchy. This is needed to store sets of values for an attribute (e.g.
“personal.spokenLanguages = (‘de’, ‘en’, ‘fr’)”) or to provide several data sets (e.g. “personal.address(1).street”, “personal.address(2).street”).

The main features in our approach can be summarized as follows:

- hierarchical attribute space
- values at any level can be sets (multiple values)
- domain specific standard set and additional application specific attributes
- special types for interests, relationships and ratings
- ontology to define attribute hierarchy, attribute names and data types

Since attributes can be set by different sources they have to store meta information about who has changed them. In addition to knowing who has set some data it might often be necessary to have a prove for this. If attributes are to be used for one application only, the service could store these attributes locally (and only store the other attributes globally in the identity management service) – but there are scenarios where attributes should be exchanged among services and still have to be trusted (e.g. attribute that user has bought for more that US$1000 at one e-commerce site which entitles her for special discounts in other services). The two solutions to this issue are that the identity management service itself guarantees the source of the data or that the data source digitally signs the data so that anybody can check the origin and the integrity. We have chosen the second possibility. The repository servers offer a possibility to sign any sub-hierarchy or subset of attributes in the repository and store the signatures.

5 Related Work

Work related to exchanging user profiles and identity management can be found both in industry and research. First, there is work on client based profile management and provision to servers via form-fill-in or P3P in infomediaries. Second, there are already several server based identity management solutions. And third, there is basic work on identity management and privacy.

5.1 Infomediaries

Today, user-related information is stored by different services independently in proprietary ways. Alternatively, user profile information can be stored on the user’s computer, and be provided to services when needed. This could lead to higher trust because personal information is located near the user and because the usage of profile information can be controlled and monitored.

Client-side user profile storage is implemented by so-called infomediaries. Infomediaries are (small) applications on the client computer, which manage user profiles and offer services such as automatic fill-in of Web forms or P3P interfaces for exchanging the information with services (Cranor 1999).

Examples for infomediaries are

- Jotter (www.jotter.com) or
- Persona (www.persona.com).
Some infomediaries have additional features for automatically sharing information with marketers of products or services they have expressed interest in.

The main problem of client-side storage of user profile information is that it is not portable (Mulligan & Schwartz 2000). Personal information stored on one computer (e.g. at work) cannot be easily transferred to another one (e.g. at home or a mobile device). An additional problem with today’s infomediaries is that the definition of access rights is possible but much too complex for everyday usage.

5.2 Single-Sign-In and Server-based User Profile Databases

While the infomediaries focus on user control of identity information (authorization) server-side solutions are often more service-platform-centered and focus on authentication. These solutions relate back to multi-server authentication solutions like Kerberos (Steiner et al. 1988). In such single-sign-in solutions different servers or services share one service to authenticate users. Often the single-sign-in solution is extended to a user profile that is shared among the different services.

Today different software vendors offer such single-sign-in solutions for Intranets. The solutions are mainly based on (X.500) directory services or at least accessible via the LDAP directory access protocol.

While the single-sign-in solutions mentioned before are tailored for Intranet usage, global solutions like DigitalMe from Novell (www.digitalme.com) or Microsoft Passport (www.passport.com) extend this approach to a service for Internet usage. The core of these services is a central user profile directory (operated by Novell or Microsoft). Users can store and maintain their personal data in these directory servers via Web interfaces. Services that are certified by the profile storage operator can get access to the authentication and profile information when a user tries to log in at these services.

The systems are very similar to what we have in mind with the IDRepository. The main difference is in the missing orientation towards the profile owner, which shows in the profile data scope, missing access right definition, and the concentration on one profile storage operator. So, the information stored in the repository is limited to “basic e-commerce information” like delivery information (name, address) and payment information. Profile owners do not have the possibility to define access rights to their needs or to get information about the profile usage.

Other central user profile repositories are even more focused on marketing. So iFAY (www.ifay.com) or Yodlee (www.yodlee.com) support clustering users and making the information about the affiliation to clusters available to services that pay for it. In addition to the large identity management networks like Microsoft Passport several smaller projects have appeared. Examples are XNS (www.xns.org) and Live-id.org (www.live-id.org). These companies mainly follow a federated approach that allows for different identity servers operated by different companies.

To allow interoperability, the identity management providers and other companies that are already operating large Internet identification services like AOL, eBay, MSN or Visa have joined in the Liberty Alliance to develop a standard for connecting their identification and user profile storage services in a federated way. See (AberdeenGroup 2002), (Sun 2002a, 2002b) for more information on the industries viewpoint on federated identity services.

The Liberty Alliance is currently on the way to define an open standard for the representation of identities, for the authentication of users and for authorizing access to user profile information. The goal is to make it easy for services that are storing user
profile information to exchange the information among each other (Liberty 2002). However, the focus of Liberty Alliance again is on the services. There is no real user control built into the proposal yet.

5.3 Profile Information Exchange

Other related work in the commercial field is about exchange and synchronization of user profile information among users or among applications of one user. Examples for the replication of user data among users are business card exchange services (see www.cardxchange.net for one example). In these services users can store their contact information (and any additional attributes) and make a subset (view) of this information explicitly available to other users. When the information is changed by the owner the electronic business card changes at all places or the people replicating the information are notified by email. Similar functionality is often built into Community Support platforms. So, various alumni platforms provide some form of address exchange functionality. An interesting generic approach in this area is the solution by XNS (www.xns.org) and OneName (www.onename.com).

Exchange and synchronization of PIM data (Personal Information Management – calendar, address and todo lists, notes) up to now has been restricted by the large number of proprietary protocols on the market, each focusing on only a small number of devices, applications and data types. SyncML (www.syncml.org) - an open industry standard for the synchronization of remote personal data across multiple networks, applications, platforms and devices - resolves this issue by providing a level of interoperability that is not possible with the industry's current proprietary synchronization protocols. In the context of SyncML also different data standards like vCard and vCal are promoted.

5.4 Identity Management and Usability

While commercial approaches concentrate on small subsets of user profile information and on service-centric implementations the research community is working on more general topics.

First we have to mention work dealing with what “identity” is and how the identity is used or determines interaction in online communication. Examples for work in this area are from Donath (1998) on Newsgroups and from Churchill and Bly (1999) on MUDs.

Security issues and legal implications are the topic of another large corps of work. Examples are (Damker et al. 1999) or (Köhntopp & Bertold 2000). The work of Köhntopp also discusses general questions of identity management and user interfaces of identity management tools.

Finally, there is some work on the usability of security tools. At the University of Freiburg for example Jendricke and co-workers are discussing identity management as a concept for achieving usable security in the internet. They focus on reducing the complexity in user interfaces and are developing a security tool called Identity-Manager that allows the user to easily choose different views and identities (Gerd tom Markotten et al. 2001, Jendricke & Gerd tom Markotten 2000).
6 Summary and Conclusions

The availability of user profile information will be important for future Internet based Electronic Commerce and Community Support services. Information about the users is needed for performing transactions, for presenting users to each other and for providing personalized services.

Identity management and central user profile repositories might help

- to motivate users making user profile information available (because they have control and awareness about who is using it)
- services to provide effective personalization without cold-start problems

These two effects could help to boost the use of personalization in online services.

However, there are still some challenges to be addressed. From the technical point of view the most important issues are:

- how to specify (and enforce) access rights (especially including usability and user interface issues)
- how to represent user profile data to make it usable by different services (up to now nobody has dealt with user profile structure very much)

Some of the issues cannot be solved through technology alone. Especially the issue of access right enforcement. As already discussed in the P3P project of the World-Wide-Web-Consortium a certification of services is needed to ensure that the services make correct statements about planned user profile usage. An issue linked with the service certification is the selection of a trusted operator for the identity management service. Here we have taken an approach that allows different providers to operate identity servers and allows the user to select.

To gain trust from the profile owners a solution has to clearly support

- definition and handling of different access rights and/or sub-identities
- provision of awareness of access to the profile information

The functionality has to be provided in an intuitive way and has to cover the emerging mobile applications that also need user profile information for performing their services. All this has to be provided by different (trusted) operators for identity management servers to choose from.

While this approach of federated identity management services is already taken up by big consortia (see section on related work), current approaches are mainly authentication centered and still too much focused on services to become global identity management solutions. In our opinion more user-centered solutions from service independent providers with a focus on access right definition (authorization) and usability are needed for the future. In this context work on making the usage and configuration of these services as intuitive as possible to the profile owner is a central requirement. Here more work is needed. Another topic where additional work is needed is on business models for future identity management providers. First ideas are drawing from analogies of “user profile banks” with classical banks that have gained trust and are providing access to money from everywhere.
References


