PersoBOX: A Personalization Engine between ERP System and Web Frontend

Christoph Adolphs
Institute for IS Research, University of Koblenz-Landau, Christoph.adolphs@uni-koblenz.de

Petra Schubert
Institute for IS Research, University of Koblenz-Landau, petra.schubert@uni-koblenz.de

Follow this and additional works at: http://aisel.aisnet.org/bled2008

Recommended Citation
http://aisel.aisnet.org/bled2008/7
Abstract
The demand for personalization functions in e-shops is increasing steadily. In order to fulfil customer requirements best and to stimulate the customer’s buying experience positively, companies are aiming at an easy technical solution to the integration of ERP master data, CRM data, and transactional data from web shops. The current paper presents the state of the art in personalization in e-commerce and summarizes remaining problems. An integrated toolset, the so called PersoBOX, is introduced as a solution which connects the realm of ERP systems with web shops. We present a schematic architecture of the PersoBOX describing the data flows, as well as processes and functions to be implemented. The presentation of the architecture is a preliminary result of an ongoing research project in the area of personalization.

Keywords: Personalization, ERP systems, CRM, Web applications

Introduction
The theme of the 21st Bled Conference is “Overcoming Boundaries Through Multi-Channel Interaction”. Multi-channel approaches involve electronic channels where customers interact with Web applications. Personalization plays an important role in this context. We need to look at two different business software systems when analysing customer self-service applications: (1) ERP systems and (2) Web applications (mainly e-shops). ERP systems are used by internal staff members in their daily routine. They are familiar with the user interface and the supported business processes. Customers, on the other hand, interact infrequently with the Web application (e-shop). The supported functions for e.g. the creation of an order process need to be as comprehensive as in the ERP system, but they also need to be self-explanatory, guiding the user through the process.

Personalization is a means of facilitating this ease of use by presenting the customer a pre-defined interface for his/her personal needs. It is an interdisciplinary topic with which research papers,
particularly at conferences on marketing and computer science, have increasingly been concerned. Personalization is targeted at fulfilling a special customer or user requirement. It can be aimed at people as well as at organizational roles in companies, such as a purchasing agent. Personalization in our understanding starts *after the login*. The process is context sensitive (regarding the output for a certain user) and requires learning (by the system).

In this paper, we present a concept for a software tool, the PersoBOX, which acts at the interface between the ERP system and the Web application. The tool has access to both sides – the internal ERP world and the external Web application. The information stored on both systems is used, processed and stored in specialised databases. The PersoBOX provides the Web frontend with the necessary information to respond to a single customer (e.g. by providing recommendations, statistics, individualised product catalogue, and quick access to the customer profile).

The research methodology which we used for our research is a mixture of action research and design research. For seven years, *action research* projects in the area of personalization have been carried out with industry partners. The results of these projects triggered the idea for the PersoBOX and helped to define the basic requirements. The PersoBOX itself follows a *design research* approach because it represents a generic software component which can be implemented between a standard off-the-shelf ERP system and a Web application.

The following chapters give an overview of the current literature in the area of personalization, and the integration between ERP systems and web shops. The technological concept of the PersoBOX is described and its application in practice is discussed. The paper concludes with an outlook on future research.

**Background and Related Work**

Deitel et al. [2001] defined personalization as using „information from tracking, mining and data analysis to customize a person’s interaction with a company’s products, services, web site and employees“. Mulvenna et al. [2000] understood personalization as „the provision to the individual of tailored products, service, information or information relating to products or services. This broad area also covers recommender systems, customization, and adaptive web sites“. Adomavicius and Tuzhilin [2005a] summarized that „personalization tailors certain offerings (such as content, services, product recommendations, communications, and e-commerce interactions) by providers (such as e-commerce web sites) to consumers (such as customers and visitors) based on knowledge about them, with certain goal(s) in mind.” These definitions imply a close relationship between personalization and the recommendation of items. Even if recommender systems are undoubtedly an interesting part of personalization, there are many other personalization functionalities that are geared to improving customer loyalty. Examples of such functions are personal shopping lists, customer-specific assortments or extensive checkout support. Therefore we decided to follow the broader definition of personalization provided by Riecken [2000]: „personalization is about building customer loyalty by building meaningful one-to-one relationships; by understanding the needs of each individual and helping satisfy a goal that efficiently and knowledgeably addresses each individual’s need in a given context.” The possibilities of personalizing the user interface were pointed out by Peppers and Rogers [1997] as well as by Allen et al. [2001].

Personalization uses information about customers. The general term used for stored customer information is the „user profile” or in the context of electronic shopping the „customer profile”. There are various ways how e-shop operators can cultivate customer profiles: „historically” by storing (1) interaction with the website (click stream), or (2) purchase transactions, or „explicitly” by asking (3) for preferences, or (4) ratings, or (5) by recording contextual information (e.g. time, date, place). What formerly seemed to be possible only for the corner shop whose storekeeper knew all the customers personally, reaches new potential in the online medium where every customer leaves traces and thus „teaches” the system how to treat him in a different way than other customers. This form of personalization becomes feasible with the use of predefined rules which can be built into e-commerce environments. These automatic personalized websites do not achieve the high quality of corner shops but they help to establish a personal dialogue with the customers.
and thus to tie them closer to the electronic offer. Additionally, the time spent by the customer to “teach” the system is assumed to lead to increased switching costs.

Over the past years there has been a lot of research in the broad field of personalization focusing on recommender systems [Sarwar et al. 2000; Adomavicius and Tuzhilin 2005b; Herlocker et al. 2004], privacy concerns [Ackermann et al. 1999; Preibusch 2005; Risch and Schubert 2005], human-computer interfaces (HCI) [Spiekermann and Paraschiv 2000; Bareis et al. 2002; Esswein et al. 2003] and personalization as a marketing approach [Peppers and Rogers 1997; Pal and Rangaswamy 2003; Schubert and Koch 2002].

In its most common form, a personalization function can be defined as the combination of profile data with personalization techniques to ease and support human-computer interaction. In this paper we focus on personalization functions.

The ability to deliver personalization depends upon (1) the acquisition of a “virtual image” of the user, (2) the availability of product meta-information and (3) the availability of methods to combine the datasets in order to derive recommendations for the customer.

The collection and use of customer information has a downside – collecting customer specific data may smack of spying and sniffing around [Gentsch 2002]. Other effects of careless data collection activities are intentional false statements which lead to bad data quality and therefore useless customer profiles [Treiblmaier and Dickinger 2005]. The importance of privacy and security aspects in the field of CRM was pointed out in a recent survey by Salomann et al [2005].

There is an ongoing discussion on privacy that is closely related to personalization. Cranor [2003] discussed privacy risks associated with personalization in e-commerce applications and provided an overview of principles and guidelines to reduce these risks. He identified the following privacy concerns: (1) unsolicited marketing, (2) system predictions are wrong (incorrect conclusions about users), (3) system predictions are too accurate (the system knows things nobody else knows about the users), (4) price discrimination, (5) unwanted revelation of personal information to other people, (6) profiles could be used in a criminal case and (7) government surveillance.

With the above mentioned information from literature in mind, we started our action research for the development of personalization functions described in Chapter 0 and we developed a generic concept for a personalization (section 0).

**Action Research: In Search of a Generic Personalization Component**

This chapter presents the results of a long-term research project in the field of “Personalization of E-Commerce Applications” funded by the public. The project involved three different universities and ten companies which jointly worked on the development of personalization issues. The authors fulfilled the role of the leading project coordinators and project managers. The activities started in the year 2000 with an empirical study about the current state of personalization in E-Commerce applications run by small and medium enterprises (SMEs) [Leimstoll and Schubert 2002]. The findings showed that although there were companies such as Amazon and eBay that effectively demonstrated how personalization works; only very few companies were able to implement similar features into their web environments. The companies that took part in the survey showed deficits in strategy as well as a lack of technical readiness in their information systems and databases. Nevertheless, most companies acknowledged the importance of a personal relationship with the customer and indicated that personalization issues would be given high priority in the future [Leimstoll and Schubert 2002; Risch 2007]. The empirical results encouraged us to approach companies which showed an interest in the development of personalized functions in their web sites.

The research method that we used for the work with the companies is action research [as e.g. described by Cunningham 1993, Baskerville and Wood-Harper 1996, Kock 1997, Lau 1997]. There are two essential aims, common to all literature on action research: to improve and involve. Our
research project on personalization was targeted at the improvement of current e-shop processes and design (improve). The employees of the partner companies and the authors were intensely involved in the design process (involve). The first functions were implemented in 2001. After that they were deployed, repeatedly tested and refined over the years with the help of the partner companies. The process was thus cyclic, participative, qualitative, and reflective. As it is typical for action research, we aimed at the following three areas: (1) improvement of practice, (2) improvement of understanding of practice by its practitioners, and (3) improvement of the situation in which practice takes place. The results of the action research are described in detail in [Schubert et al. 2006].

Early in the project we decided to focus on integrated systems where an e-shop component would run as an extension of an ERP system. The objective was to reuse the existing ERP functionality for the selling process (so far used by the sales people) and to adapt it to the customer interface. The requirements of the partner companies made it evident that stand-alone webshop environments without a connection to an existing ERP system would not work for companies in the long run. In the current project, we went one step further and detached the personalization component from both the ERP system and the webshop by constructing a mediating platform, the so called "PersoBOX".

### Design Research: Architecture of the PersoBOX

In our research, we followed the steps of design research as presented by Takeda et al [1990] and Vaishnavi and Kuechler [2006]. Following these authors, design research consists of five major steps: (1) Awareness of Problem, (2) Suggestion, (3) Development, (4) Evaluation, and (5) Conclusion. The first step of the design process is the awareness and description of the problem which has been addressed in Chapter 0. The problem in hand is the creation of a standardized way to add personalized functions to an e-shop out of the context of ERP system data. If, for example, a company with an integrated ERP system wants to present personalized product offers to their customers while filling the shopping cart, perhaps by using a recommender system, then the implementation and customizations will need to be entered manually in most existing e-shops.

In order to determine the results of different personalization functions, it is necessary to be aware of the required inputs for each function to be applied. Groups of input types (group members have equal semantics) often do not have an equal syntax within the group. Therefore there has to be a pre-process which transforms the different types of data within their varying syntax into a normalized and unified state. Usually, this transformation is done with the help of matching tables. If no matching pattern is found for the data, a mechanism will be needed for the manual population of a transformation rule.

In our action research, we identified more than 120 personalization functions which we needed to consider in the PersoBOX. Each of them can be implemented in various ways. In order to determine the correct technical implementation for the designated output platform, there is a need for

- clearly maintaining the participants' preferences
- output specific choice of suitable functions
- dynamic combination of functions with data and preferences and
- output dependant deployment of special functions.

There is also an overall need for giving the user the possibility to maintain personal settings. These are determined by explicit settings contributed by the user on the one hand and implicit settings generated by the system on the other hand.

Additionally, the PersoBOX allows the customer to check whether explicit settings of implicit recommendations concerning him are suitable. He may even define rules to be used upon his profile in order to receive only desired personalized function outputs. Within these settings the customer can put himself into a defined group, to separate for example personal from business-related data and their utilization as well as the utilization of functions.
The focus of the remainder of the paper is an overview of the architecture of the PersoBOX. This reflects step two of the design research approach introduced in the first part of this chapter. Steps three, four and five are the focus of a current research project and will be described in a future paper.

**PersoBOX: Request View**

The development of a software architecture connecting the complex worlds of ERP systems and e-shops requires a fundamental model of all participants (including software systems) showing the interaction between software systems, people and other possible instances.

A simple model of the PersoBox was created by using only four participating instances. First, there is the group of **ERP system operators**. Within the enterprise boundaries there are several pieces of information concerning customer and product details. Apart from the basic (static) data, there is sometimes additional data available which can serve as an input for further personalization functions (e.g. CRM systems and data warehouses).

Second there is the group of **e-shop operators**. Transaction data and user behaviour can be determined within specific analysis methods (e.g. clickstream analysis and purchase correlations between consumer groups or a simple log file analysis).

Third there are the **customers** who buy goods and services in the e-shop. The needs of these participating users are in the spotlight of the approach described in this paper because this group will receive the output of the personalized functions.

At last there is the **operator of the PersoBOX** who hosts the platform as a service for the other parties.

Separating four groups of participants does not necessarily mean that there are four different people or organisations involved. If for example the organisation running the EPR system decides to host the e-shop as well there will be only three instances of the four groups in our model. If the enterprise in our example is also responsible for the PersoBOX there will be only two instances of our group model.

In the next section we present a brief overview of the general mechanisms that are planned to be supported by the PersoBOX. The purpose of the software system is described from the consumer’s perspective with a request-answer model. In this model an exemplary consumer of a service offered by the e-shop is displayed within his unique electronic purchasing process.

**Figure 1: Request-answer view**
The request process is shown in Figure 1. The figure shows two mainly independent sections. The first section begins with a collection of data made available (which is) based on a customer’s output preference prediction. These predicted settings influence the data stored in the PersoBOX. But even without any prediction (empty prediction), data such as product details and grouping criteria can be transferred to the storage entity of the PersoBOX.

The second section deals with output creation. Whenever a customer opens a page on the website, an individual data set from the data store is used to generate a personalized output.

The customer’s request is analysed, suitable personalization functions are chosen, and the contents of the data store are dynamically transformed into abstract output objects. These output objects are deployed in a suitable piece of program (dynamically created code being compiled at runtime).

After this step the customer may fill his shopping cart.

**PersoBOX: Business View**

Figure 2 displays the business view of the PersoBOX using the eXperience methodology as described by Schubert and Wölfle [Schubert and Wölfle 2007]. In contrast to the data-oriented view presented in Figure 1, Figure 2 shows the scenario and the process timeline of the participants from a business point of view.

For a detailed and complete view of the architecture of the PersoBOX it is important to identify each of the owners of the sub-processes. The connections between the players indicate the interfaces and processes that are shown in the next section.

![Figure 2: Business View following the eXperience methodology [Schubert and Wölfle 2007]](image-url)
HTML frameset with reserved spacing and linking to external web content). The e-shop requests personalized functionality from the PersoBOX for those areas that have been defined to be personalized for the user. The PersoBOX itself has access to its own unified database or may retrieve specific parameters from the e-shop or the associated ERP system on request.

After preparing the necessary data and the required functions, the referrer (URL of the function) or a complete static tool (e.g. compiled Java code or the output from web scripts like php in HTML) is transferred to the e-shop to be assembled and presented to the customer in the following step. The requests in the process of page creation by the e-shop to the PersoBOX can be omitted if a set of prearranged personalized functions is deployed by the PersoBOX. In this way, the calculating process at runtime can be reduced to a minimum with the help of the data that has already been collected at the time of output prediction. In case no pre-personalized functions can be retrieved, the search process may be started on the fly to update the deployed function data in real-time.

**PersoBOX: The Architecture**

Figure 4 displays a more detailed view of the architecture of the PersoBOX. The PersoBOX makes use of several inputs and outputs that work as connectors to its environment. We already described these interfaces in the data-oriented request view.

The four phases shown in the architecture view are called input, input processing, output processing, and output. They were developed in accordance with the Customer Profile Life Cycle by Schubert and Leimstoll [2002] shown in Figure 3. Step 1 (modelling profiles) is only performed once for the PersoBOX (in the initial customization phase).

![Customer Profile Life Cycle](image)

*Figure 3: Customer Profile Life Cycle [following Schubert and Leimstoll 2002, 21]*

Modelling the profile is done by unifying the input data according to translation tables (matching tables). The three remaining phases are repeated twice in the personalization process.

The final output is generated in an asynchronous way. First, the input factors are processed. Afterwards, an output is generated and stored in a database for later use. This completes the first cycle of the Customer Profile Life Cycle. When using the output of the first cycle as input for the second one, a new round is started. Processing the data (generating the e-shop output) produces personalized functions to be used by the e-shop operator.

For the design of a complete architectural view of the PersoBOX, we believe that it is important to divide the inputs of e-shops and ERP systems into information schemes which describe the structure and syntax of the data and raw data. The resulting scheme connector enables the PersoBOX to process the data and add semantics to it.
The e-shop operator and the ERP system operator have a direct influence on the processing of the data. They provide structural and functional limitations and prerequisites to be fulfilled by the PersoBOX. This is done in a preliminary step which we call system customization by utilizing a so-called “input profile system”. The influence of the two parties is modelled in Figure 4.

Figure 4: Architecture of the PersoBOX
The PersoBOX needs some basic user profiles in order to operate properly. For a new customer, a set of preferences (cf. Figure 4 - G) is either entered explicitly by the customer or implicitly compiled from third party data (cf. Figure 4 – F.4). The profile is constantly improved during later customer interactions with the web site. Possible inputs are the click stream and buying behaviour which are analysed and transformed for use in personalization functions (e.g. an analysis of transaction data to identify whether a specific site user buys electronic products or not).

The input connector provides an explicit inward flow of settings to the customer preferences profile.

Risch et al. [2006] presented a review of possible personalization functions including a high level overview of possible input databases (cf. Figure 5). The input sources recommended for the PersoBOX are similar to the ones presented in their model but structured and extended in a more technical way to facilitate later implementation.

![Figure 5: A technical view on personalization [Risch et al. 2006]](image-url)

The input processing in the PersoBOX approach is based on the generic layout from Risch et al. [2006] (layer 2 and 3). The ETL Process ("(E)xtracted from the original source, (T)ransformed into a pre-defined structure and finally (L)oaded into a centralized data warehouse for customer profiles" [Risch et al. 2006, p. 4]) used in this model does not suffice to be adopted directly in our
PersoBOX: A Personalization Engine between ERP System and Web Frontend

approach because our model requires more detailed knowledge about single process steps (e.g. we separated the transformation phase into unification and filtering).

The results of the customization process and the basic settings of the PersoBOX are stored in an internal database. We envision a process which provides passive and active services for other components in the PersoBOX in a way of a Service Oriented Architecture (SOA). With this kind of approach all other components of the PersoBOX using the interfaces can easily be integrated into the system even if they are designed and programmed by third parties. The process responsible for providing internal interfaces is the **Output Forecaster** (cf. Figure 4 - A) which can be asked for pre-customized restrictions and requirements. The Output Forecaster is also in charge of setting basic filtering rules to eliminate useless data from the input stream. This filtering is done by a rule-based filtering engine with saving routines to store the data afterwards.

Before any data can be processed by the filtering mechanism (cf. Figure 4 - B), the amount of data needs to be converted into a unified representation of the data (**unification process**, cf. Figure 4 - C). This step eliminates undesired differences in namespaces and simplifies the usually heterogeneous syntax of different input sources. The responsible process is pattern-based and operates on input dependent matching tables.

The triggering of the data transfer can be initialized by a pull operation from inside of the PersoBOX or by a push operation in one of the systems of the participants (ERP system operator, e-shop operator or customer). To represent both possibilities we introduced a process called the **Data Input Scheduler** (cf. Figure 4 - P). This scheduler is able to initialize data transfers by fulfilling a set of internal rules (e.g. trigger a special transfer at a specific time) and pulling the required amount of data, or by being called by participant processes that announce new transmissions to be received soon.

Because of the dualistic character of this interface process (the process possesses an input and output role), it should have connections on the input and output side of the PersoBOX. Our approach allows it to categorize the scheduler as an input process because its only purpose is to keep up a data flow inside the system. It is mainly input related.

Both the phase of input processing (customizing, unifying, filtering, scheduling, and storing) and the phase of output processing are separated from each other. Data collection can be performed before any user request is sent to the e-shop. In the event of missing data in the output processing, there is the possibility of interacting with the input processing again to trigger the retrieval of additional data.

The starting event of the output processing is a customer requesting a page in the e-shop. This request is computed by a middleware engine of the e-shop, manipulated, and transferred to the PersoBOX. The **Request Scheduler** (cf. Figure 4 - O) triggers a process called the **Function Chooser** (cf. Figure 4 - H) after updating the **Output Forecaster** with additional information. The Output Forecaster is the connection to the phase of input processing.

The Function Chooser checks for limitations, the Output Forecaster sets and afterwards combines these recommendations with the relevant customer preferences. After that, the Chooser process is ready to choose a set of personalization functions out of a pool of possibilities.

As mentioned above, Risch et al. [2006] developed a personalization map with a detailed collection of functions for personalization processes. These functions are implemented in a generic form and stored in a database which is called up by the Function Chooser. The result-set is transmitted to the **Application Generating Process** (cf. Figure 4 - J). This process receives the functions which should be deployed from the Function Chooser and determines a concrete implementation strategy and output format from the updated Output Forecaster. The **Application Generator** transfers the program code to a **Deployment Process** (cf. Figure 4 - M) and triggers a **Data Transfer Scheduler** (cf. Figure 4 - K).
The Data Transfer Scheduler starts the transfer of data required by the personalization functions to a user-specific database.

The Application Generator also checks whether the data, to which the Transfer Process has access, is sufficient for all implemented personalization functions. If any of the required data is still missing, a request for data retrieval will be sent to the input processing.

The Deployment Process is able to connect the implemented functions to the collected personalized data and creates an output which the e-shop can process afterwards. In addition, the PersoBOX provides a customer interface which displays the customer’s preferences.

The described processes and databases are the basis of the overall functionality of the PersoBOX and serve as the blueprint for the current technical development.

Conclusions and Future Research

This paper presents the preliminary results of an ongoing research project in the field of personalization. In previous steps, personalization functions had been collected and requirements for a personalization toolbox, the PersoBOX, had been identified. The Customer Profile Life Cycle [Schubert and Leimstoll 2002] and the technical model on personalization developed by Risch et al. [2006] were used as a basic template for the architecture of the PersoBOX.

Future research will address an array of open questions. There is a need for looking closer at the personalization processes and pre-processes in each of the four phases in the architectural view of the PersoBOX. Within the input and the input processing phase there is more than one possible strategy to unify data according to customized rules. Additional strategies include generic algorithms and neural networks as solutions for rule-based approaches. Other challenges for the PersoBOX itself and for future research approaches will be: adding semantics to already processed input data, working with these semantics, and developing dynamic mechanisms to connect functions, data and semantics.

The impact of third party input data suppliers (community portals such as “facebook”, “fellow-web”, “wer-kennt-wen” or “openBC/Xing”) to the functionality of the PersoBOX also seem to be an interesting challenge for additional research steps. The actual tool described in this paper is currently being developed.

References


