

Challenges and Benefits of the Semantic Web for User Modelling

Peter Dolog and Wolfgang Nejdl

Learning Lab Lower Saxony
University of Hannover
Expo Plaza 1, 30539 Hannover, Germany,
{dolog, nejdl}@learninglab.de,
<http://www.learninglab.de/~dolog>,
<http://www.kbs.uni-hannover.de/~nejdl>

Abstract. The aim of this paper is to discuss how distributed learner modelling can benefit from semantic web technologies and which challenges have to be solved in this new environment. Heterogeneity of personalization techniques and their needs raise the question whether we can agree on one common data model for user profiles, which supports these techniques. In this paper we discuss an approach where a learner model can be distributed and can reflect features taken from several standards for a learner modelling. These features can be combined according to the requirements of specific personalization techniques, which can be provided as personalization services in a P2P learning network. RDF and RDFS as key tools of the semantic web allow us to handle such situations. We also sketch an architecture for such a network, where this approach can be realized.

1 Introduction

Internet as an open environment provides us with the opportunity to share and reuse resources already available. Heterogeneity of users and resources in the web stresses the importance of customized (personalized) delivery of the resources.

A wide range of personalization techniques has been introduced based on metadata about a user or learner. The first category of techniques are based mostly on adapting user interfaces, navigation and content selection and presentation according to the user's performance in a particular domain. The performance is often evaluated in a small closed domain, e.g. an electronic course at the open university. These techniques are usually called *Adaptive hypermedia techniques* [2].

Another type of techniques is based on interests, preferences, likes, dislikes, and goals a user has. This information is mostly stored at some kind of modelling server [6]. These are the so-called *filtering and recommendation techniques*. They recommend resources according to features extracted from a resource content or according to ratings of a user or learner of similar profile.

As it was already pointed for example in [10], such single user model, which stores everything in one model on one place has its disadvantages. A user model designed for experimentation with a specific technique usually reflects the specific requirements

of that technique. This means that the schemas designed for different personalization techniques do not only differ syntactically (schema elements used to encode user information) but also semantically (i.e. that elements used for representing certain user feature can have either different syntax and different meaning or the same syntax and the different meaning). They can also differ in structuring the user's or learner's feature space.

There have been attempts to standardize a learner profile (see e.g. IEEE Personal and Private Information (PAPI) [4] or IMS Learner Information Package (LIP) [5]). These standards have been developed from different points of view. The PAPI standard reflects ideas from intelligent tutoring systems where the performance information is considered as the most important information about a learner. The PAPI standard also stresses the importance of inter-personal relationships discussed also in [10]. On the other hand the LIP standard is based on the classical notion of a CV and inter-personal relationships are not considered at all.

In this paper we report on our experience in developing the first version of a learner profile to support simple personalization techniques in the ELENA project. The aim of the project is to integrate heterogeneous learning resource and service providers and enable personalized access, use and delivery of the resources and services bound to them. The network is based on the Edutella P2P infrastructure [8]. This work is the extension of our previously published work on integrating adaptive hypermedia techniques into open RDF based environments [3].

The rest of the paper is structured as follows. First, we motivate our work by discussing some characteristics of open environments (section 2). We then remark on how users or learners can be modelled by utilizing RDF and RDFS in section 2. We then discuss our approach for developing a first version of a user profile for personalization services to support personalized search in the EU/IST ELENA project in section 5 and discuss its relationship to existing user profile standards as well as its embedding into a P2P architecture.

2 Some characteristics of open learning environments

There are several characteristics of open learning environments integrating heterogeneous resource providers, which distinguish open environments from most other currently studied systems. First of all, the resource can appear and disappear in ad-hoc manner. Peers, which provide such resources can appear and disappear randomly.

Resources are authored by different people with different goals, background, domain expertise and so on. Providers of the resources can maintain the resource in proprietary databases. They can already have some personalization techniques implemented for purposes of their context. The user or learner model usually reflects the context of the techniques as well. The resources are accessed and consumed by people which differ in a wide range of characteristics.

User or learner features can already be maintained in some of integrated systems. *Human resource management systems* as one of the modules of enterprise resource planning systems incorporate a module related to maintaining information about employees. This module usually maintains information about employee identification,

skills, previous jobs, previous training, current role and position within company, transcripts and so on. *Task management systems* like project management systems, Sales Force Automation, or simply Outlook with its calendar contain usually information about tasks daily performed. *User modelling servers* maintain characteristics such as likes and dislikes, interests, bought products, and so on. *Electronic course providers* already have a model of learner performance. This basically refers to learner's attended courses with ratings of his performance. The performance record also refers to the formal certificates earned from the course providers and results of work or additional material used during the course. The providers can also maintain user interests, preferences and goals.

It is obvious that in such environments the agreement on common structures and scope of information about a user can be made only within specific communities or for specific personalization techniques.

3 Standards for Learner Modelling

The two most important examples for such standards are PAPI [4] and IMS LIPS [5]. Both standards deal with several categories for information about a learner. Figure 1 depicts the conceptual view of PAPI profile.

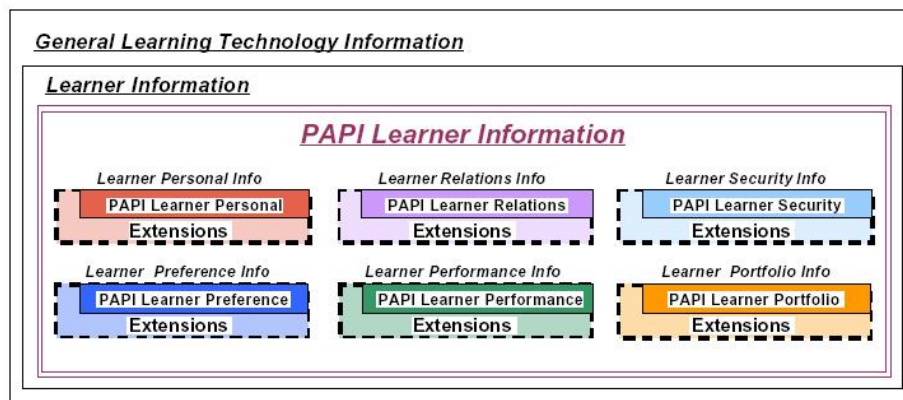


Fig. 1. The core categories in the PAPI profile [4].

PAPI distinguishes *personal*, *relations*, *security*, *preference*, *performance*, and *portfolio* information. The *personal* category contains information about names, contacts and addresses of a learner. *Relations* serve as a category for relationships of a specific learner to other persons (e.g. classmate, teacheris, teacherof, instructoris, instructorof, belongsto, belongswith). *Security* aims to provide slots for credentials and access rights. *Preference* indicates the types of devices and objects, which the learner is able to recognize. *Performance* is for storing information about measured performance of a learner

through learning material (i.e. what does a learner knows). *Portfolio* is for accessing previous experience of a user. Each category can be extended.

Similarly the IMS LIP standard contains several categories for data about a user. The categories are depicted in the fig. 2.

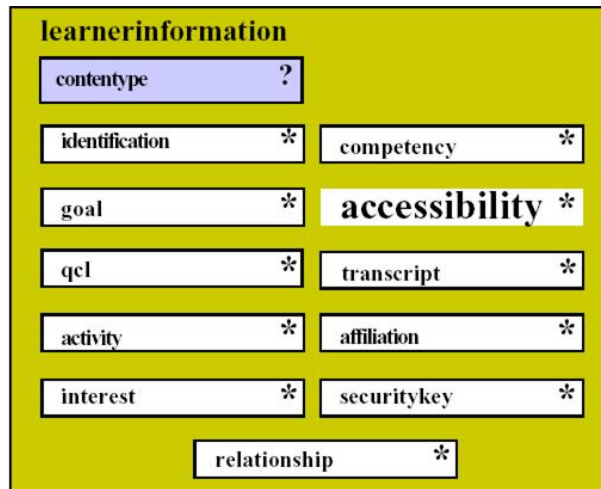


Fig. 2. The core categories for learner data in IMS LIP [5].

The *identification* category represents demographic and biographic data about a learner. The *goal* category represents learning, career and other objectives of a learner. The *QCL* category is used for identification of qualifications, certifications, and licenses from recognized authorities. The *activity* category can contain any learning related activity in any state of completion. The *interest* category can be any information describing hobbies and recreational activities. The *relationship* category aims for relationships between core data elements. The *competency* category serves as slot for skills, experience and knowledge acquired. The *accessibility* category aims for general accessibility to learner information by means of language capabilities, disabilities, eligibility, and learning preferences. The *transcript* category represents institutionally-based summary of academic achievements. The *affiliation* category represents information records about membership in professional organizations. The *security key* is for setting passwords and keys assigned to a learner.

4 Semantic web technologies

Semantic web technologies like the Resource Description Format (RDF) [7] or RDF schema (RDFS) [1] provide us with interesting possibilities. RDF models are used to describe learning resources. The RDF bindings of Learning Object Metadata (LOM) [9] can be used for these purposes. The RDF model can be used for learner description as well. RDF schemas serve to define vocabularies for metadata records in an RDF file.

There is no restriction on the use of different schemas together in one RDF file or RDF model. The schema identification comes with attributes being used from that schema so backward dereferencing is again easily possible. For example the RDF model about a learner can use an attribute `performance_coding` from PAPI standard. Field from PAPI standard can be prescribed by PAPI schema. An Instance of such a record can then look as follows:

```
<rdf:Description rdf:ID="record_1">
...
  <papi:performance_coding
    xmlns:papi="http://learninglab.de/papi#">
    number
  </papi:performance_coding>
</rdf:description>
```

The abbreviated syntax can be also used where schemas are identified directly in RDF element as a root element of the RDF description. If we want to use different elements from different schemas the syntax will be as follows:

```
<rdf:Description rdf:ID="record_1">
...
  <papi:performance_coding
    xmlns:papi="http://learninglab.de/papi#">
    number
  </papi:performance_coding>
...
</rdf:description>
<rdf:Description rdf:ID="record_2">
...
  <ims:language_preference
    xmlns:papi="http://learninglab.de/ims#">
    english
  </ims:language_preference>
...
</rdf:description>
```

It says that we are using `ims` schema element `language_preference` for encoding learner language preferences. In this case the preference is English. This element was taken from the IMS schema.

The Edutella [8] P2P infrastructure allow us to connect peers which provides metadata about resources described in RDF. Edutella also provides us with a powerful Datalog based query language, RDF-QEL. The query can be formulated in RDF format as well, and it can reference several schemas. Furthermore, Edutella is able to process insert/update requests.

An example for a simple query (in Datalog syntax) over a learner model can be as follows:

```
query(student1, X, Z) :-
performance(student1, X),
```

```
performance_value(X, Z).
```

```
?- query(student1, X, Z, V)
```

The query tries to find performance values for a student. Variable X is bound to the performance record identifiers, variable Z to performance values. However, if we query the RDF model from our previous example we will not receive any result.

If we want to update the RDF description from our example with a performance value, an insert request should be used. The statement for inserting the performance value will contain `record.1` resource as a subject, `performance_value` as a predicate and the specific value for learner performance as object.

The resulting RDF description will then contain the performance value encoded similarly to the performance coding.

5 Learner profile development

We are working on personalization in the context of the EU/IST project Elena. The aim of this project is to demonstrate the feasibility of smart learning spaces. These spaces will be realized by setting up a network where heterogeneous services and resources are provided from different peers. The provision, search, booking, and delivery services should be personalized. The first goal is to support simple personalization techniques for search service. As we mentioned, personalization techniques depend on information about a user. The question for the development of such personal profile in the open heterogeneous environment was whether to use one specific standard, combine existing standards according to our needs, to develop our own profile schema, or some intermediate solution.

5.1 Methodology employed

To answer these questions, we employed a scenario based methodology for analysing the requirements for our personal profiles. The methodology consists of the following steps:

- Collecting scenarios
- Analyzing scenarios
- Creating exemplar conceptual views of scenarios
- Abstracting and generalizing exemplar views - creating classes of user features
- Analyzing available standards
- Mapping classes from class based conceptual model to categories from supporting standards
- Refining scenarios

We could have also driven the development of a user profile by a specific subset of personalization techniques but we decided to have a broader scope taken from scenarios collected. This enabled us to analyze standards for learner profiles from different perspectives described by scenarios. Personalization techniques can be suggested according to the derived requirements from scenarios.

5.2 Scenarios analysis

We collected scenarios from several partners in the ELENA project consortium. Scenario driven approaches are quite powerful techniques for expressing high level requirements for queries user can formulate when looking for educational material or information. The scenarios helped us to better shape the diversity of personalization techniques and learner features needed for those techniques. The scenarios we analyzed are simple stories which describe usually different aspects of personalization at different levels of detail and abstraction.

One group of scenarios was oriented towards personalization based on previous experiences or performance within a domain such as finishing a certification program, enhancing the level of knowledge a learner had in some domain, and so on. A second group of scenarios was based on activities the user is involved in his job, e.g. a sale assistant should learn features of a new version of a product he should sell. He should also learn new selling techniques which are needed for this purpose. The manager wants to fill a time gap between meetings and learn something relevant to the next meeting, which can help him to succeed in negotiation, and so on.

Another group of scenarios reflected the interest to learn something from a domain which is not related to work ambitions, e.g. learn something about the history of the country the learner lives in. Interesting scenarios were also those which explicitly considered the distributed nature of the ELENA network and the availability of reusable resources, e.g. a lecturer looking for a resource to enhance his e-course or resources, to help as additional resources for newly created course. Some other scenarios reflected goals such as to move to a higher position. It means that the courses attended by people already working at the position considered as a goal can be analyzed.

An example scenario looked like the following: Bob has a meeting next month in Munich. His personal learning agent (PLA) knows Bob's activities from his calendar, and proactively tries to find out if there are any seminars that are organized in Munich the days after the meeting. A query for learning services related to computer science (Bob's preference) gives the following results: an introductory two day course in basic computer security, a refreshment course in computer security, and two seminars on advanced security technologies in networking.

Bob's PLA knows that Bob already attended an on-line course on basic security from his local university, as part of a larger seminar on computer networks, so it suggests to him only the other three seminars: one to refresh the previously obtained knowledge, and the others two to gain some new information and knowledge. One of these two advanced seminars is part of a series of seminars that lead to a certified security professional (CSP) title. Bob's PLA knows that one of the Bob's goals is to become a CSP in the future, so it emphasizes this information.

Since Bob has not forgotten much about basics in computer security yet, he decides for the advanced seminar that may help him to achieve his goal. A prerequisite for attending the seminar is knowledge about basic security. During the booking procedure Bob's PLA sends together with registration information also a certificate that confirms Bob's attendance to an on-line course on basic security. The certificate, which contains major topics that were covered in the seminar, also contains grades that are privacy sensitive. Bob's PLA decides to cover this information before sending the registration to

the seminar provider. The PLA also knows that Bob is a member of IEEE, and he is thus eligible for a seminar fee discount. After the Munich seminar Bob receives a certificate which can later be used in other seminars of the series, enabling him to become a CSP. The certificate is stored in his PLA.

5.3 Resulting Model

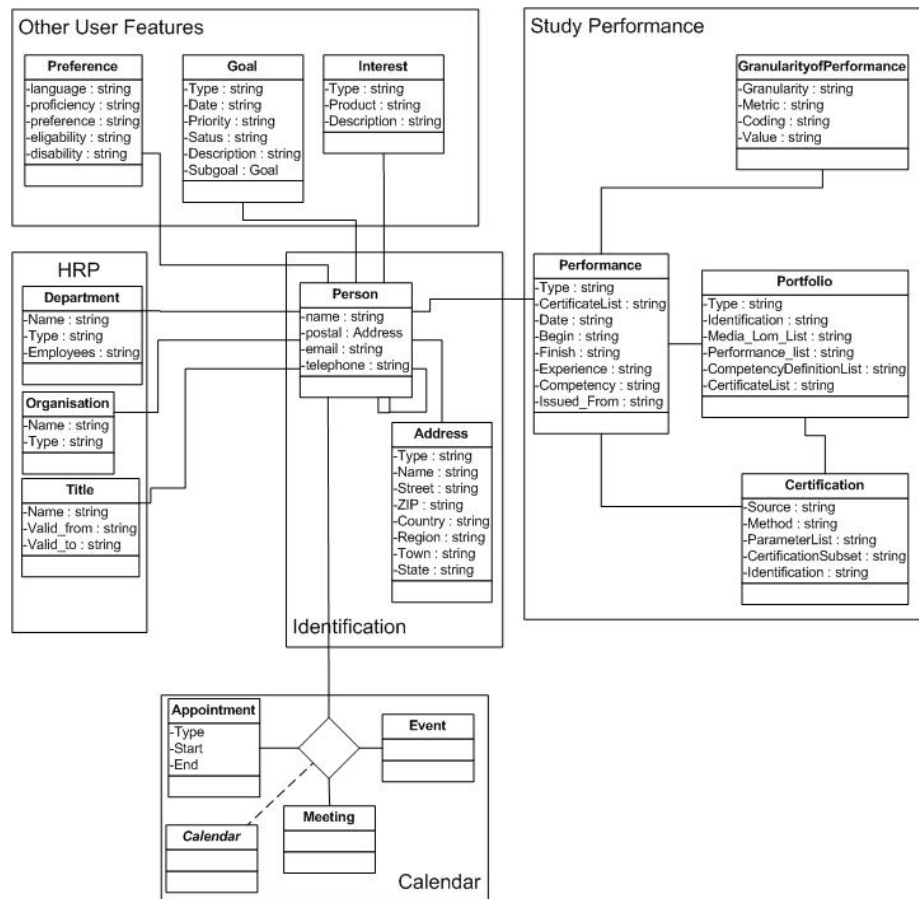


Fig. 3. First version of conceptual model of user profile for Elena.

Based on the standards before and our scenarios we proposed a learner model displayed in figure 3 as a base for further research. Because PAPI does not sufficiently deal with user features such as goal or interests, we employed a part of IMS profile for maintaining learner goals, interests and preferences. Human resource planning (HRP), calendar, and identification categories are the subject of existing systems and can be

extracted from them. Thus we do not explicitly deal with them. The identification category is needed just for internal identification. In case of the identification category we included a subset of PAPI profile field as an example.

5.4 Example

The Bob scenario mentioned in section 5.2 can be refined as follows. Bob is a `Person`, and has a working address, which represents a company he works for. This can be an instance of `Address` class from the learner profile. Bob has proficiency preference for Computer Science (`Preference` instance). He has also the goal to earn a Certified Security Professional certificate (`Goal` instance). This goal is partially fulfilled by his study performance, namely basic security course and later after attending the course in Munich on Advanced Security Technologies in Networking II (two instances of `Performance` class). Certificates from these courses can be instantiated by `Certification` class. Appointment for Munich meeting can be instance of `Appointment` class.

For space limitation we provide here just an example of an RDF instance for the performance record for the Advanced Security Technologies for Networking II course. The example is as follows:

```
<rdf:Description rdf:ID="BOB">
  <papi:performance>
    ...
    <rdf:Description rdf:ID=
      "Advanced Security Technologies for Networking II">
      <papi:issued_from_identifier rdf:resource=
        "http://www.elena.../Test_AST345"/>
      <papi:learning_competency rdf:resource=
        "http://www.kbs.uni-hannover.de/Uli/
        ACM_CCS.rdf\# C.2.0.2"/>
      <papi:learning_experience_identifier
        rdf:resource="http://www.elena.../AST34.pdf"/>
      <papi:granularity>topic</papi:granularity>
      <papi:performance_coding>number
      </papi:performance_coding>
      <papi:performance_metric>0-1
      </papi:performance_metric>
      <papi:performance_value>0.9
      </papi:performance_value>
    </rdf:Description>
  </papi:performance>
</rdf:Description>
```

6 Discussion

Our resulting learner model is based on subsets of both mentioned standards. These standards reflect different perspectives. IMS LIP provides us with richer structures and

various aspects. The categories are rather independent and the relationships between different records which instantiate different categories can be accomplished via the instances of the relationships category of the LIP standard. The structure of the IMS LIP standard was derived from best practices in writing CV's. The IMS standard does not consider explicitly relations to other people but they can be represented by relationships between different records of the identification category. However, accessibility policies to the data about different learner are not defined.

PAPI on the other hand has been developed from the perspective of a learner performance during his study. The main categories are thus performance, portfolio, certificates and relations to other people (class mate, teacher and so on). This overlaps with the IMS activity category. However, the IMS LIP defines activity category as a slot for any activity somehow related to a learner. To reflect this, IMS activity involves fields, which are related more to information required from management perspectives than from personalization based on level of knowledge. This can be solved in PAPI by introducing extensions and type of performance or by considering activity at the portfolio level, because any portfolio item is the result of some activity related to learning. IMS QCL category is similar to the certificate list enclosed in the PAPI performance and portfolio category. PAPI also has a preference category, which can be used for storing preferences about devices used for learning. This is similar to IMS accessibility category. IMS accessibility category in addition distinguishes several types of preferences, not only device preference.

PAPI do not cover the goal category at all, which can be used for recommendation and filtering techniques. PAPI do not deal with transcript category explicitly as well. IMS LIP defines transcript as a record that is used to provide an institutionally-based summary of academic achievements. In PAPI, portfolio can be used, which will refer to an external document where the transcript is stored. Although transcript is important information, it is mostly unstructured and of different format and thus hardly applicable to personalized global search in its current form. The competence category does not figure explicitly in PAPI as well. The learning experience field of the performance category can be used for encoding the competence acquired during learning.

The RDF features mentioned in section 4 have interesting implications for learner modelling. They allow us to use schema elements of both standards and also elements of other schemas. The RDF models can be accessible by different peers and even more different peers can have own representation of opposite peer. This was already discussed in [10] in the context of distributed learner modelling.

Personal learning assistants can be considered also as peers which represent a user or a learner. The metadata about a user can be provided to other peers for computing purposes (when the user allows this sharing of profiles). These peers can send and receive messages when a common language is used. Queries can be forwarded to one or more peers.

Based on this, Figure 4 depicts our current ELENA architecture. Circles represents simple providers without reasoning capabilities. Rectangles represent peers, which are able to perform programs. Multiple rectangle symbols represent metadata. Learning resources are provided through resource provider peers. Resources can be referenced by courses which represent simple learning services. Courses can be personalized by

adaptation services provided by personalization peers. Courses and resources can be recommended by recommendation services or can be filtered by filtering services. Personal learning assistants support learners or user to use the network.

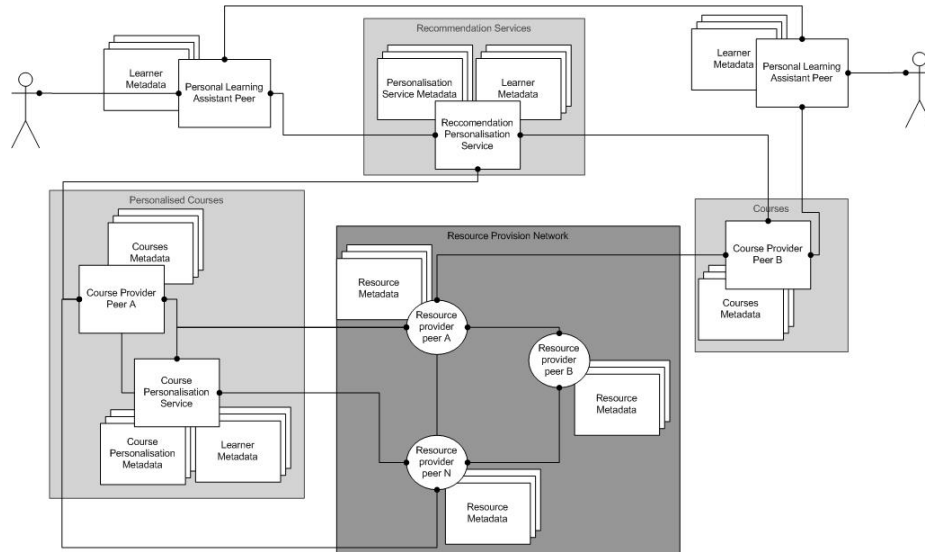


Fig. 4. ELENA architecture for personalization services.

The Datalog-based RDF-QEL language allow us to implement reasoning capabilities by using prolog or other Datalog-based systems. These reasoning can be run on specific peers when broadcasted messages are received or after receiving results of submitted query. In [3] we already described some reasoning capabilities for some adaptive hypermedia techniques for electronic course. The basic idea is to match learner performance descriptions to requirements (accessibility restrictions) a resource has for specific learner. It means that both (learning resource and learner model) use one schema according to the requirements for adaptation based on learner performance within the domain. The performance schema was derived from PAPI standard.

Aggregated user profile of users who visited a resource or liked the resource can be modelled as accessibility restrictions as well. However, different features and structure of this constraint will be needed. This model can be matched to other users searching for resources according to their goals, interests, etc. This mechanism is also used in some collaborative filtering techniques and can be represented by reasoning programs at different peers.

7 Conclusions and Further Work

This paper reported on recent work for the development of learner profile for the ELENA project. Work is not yet finished work but we can already identify features and charac-

teristics important for open environments. First of all, we can definitely benefit from learner profiles standards by combining them. The scenario driven schema development helped us to better describe and analyze our problem. Diversity of learner aspects covered by scenarios allowed us to analyze standards for learner profile from an ordinary learner perspective, who is not concerned with technical issues. Our case study for personalized search services showed that even for these restricted personalization functionality we have to build on two different standards.

Many issues still have to be resolved. The technical infrastructure for this approach to personalization has to be investigated in more detail and mechanisms for provision, searching, and using such personalization services have to be introduced. Mapping or mediating between different schemas should be investigated as well when we want to provide communication between different peers. Scenarios collected helped us in the first phase but as any user input they are at different levels of detail, precision and completeness. Thus evaluation and experiments with different queries are needed to improve our scenarios and suggested personal profile. Different strategies for employing and integrating reasoning capabilities into peers will also be investigated in the future.

Acknowledgements. We would like to thank Elena project consortium members for contribution to scenarios. This work is partially supported by EU/IST ELENA project IST-2001-37264.

References

- [1] D. Brickley and R. V. Guha. W3c resource description framework (rdf) model and syntax specification. <http://www.w3.org/TR/1998/WD-rdf-schema/>. Accessed on October 25, 2002.
- [2] Peter Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–100, 2001.
- [3] Peter Dolog, Rita Gavriloaie, Wolfgang Nejdl, and Jan Brase. Integrating adaptive hypermedia techniques and open rdf-based environments. In *Proc. of 12th International World Wide Web Conference*, Budapest, Hungary, May 2003.
- [4] IEEE. IEEE P1484.2/D7, 2000-11-28. draft standard for learning technology. public and private information (papi) for learners (papi learner). Available at: <http://ltsc.ieee.org/wg2/>. Accessed on October 25, 2002.
- [5] IMS. IMS learner information package specification. Available at: <http://www.imsproject.org/profiles/index.cfm>. Accessed on October 25, 2002.
- [6] Alfred Kobsa. Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(49):49–63, 2001.
- [7] O. Lassila and R.R. Swick. W3c resource description framework (rdf) model and syntax specification. Available at: <http://www.w3.org/TR/REC-rdfsyntax/>. Accessed on October 25, 2002.
- [8] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmr, and T. Risch. EDUTELLA: a P2P Networking Infrastructure based on RDF. In *In Proc. of 11th World Wide Web Conference*, Hawaii, USA, May 2002.
- [9] M. Nilsson. Ims metadata rdf binding guide. <http://kmr.nada.kth.se/el/ims/metadata.html>, May 2001.
- [10] Julita Vassileva, Gordon McCalla, and Jim Greer. Multi-agent multi-user modelling in I-Help. *User Modeling and User-Adapted Interaction*, 2002.