Harmonization of heterogeneous learners profiles

Carole Eyssautier-Bavay, Stéphanie Jean-Daubias, Jean-Philippe Pernin

Harmonization of heterogeneous learners profiles

Carole Eyssautier-Bavay[1], Stéphanie Jean-Daubias[1], Jean-Philippe Pernin[2], Marie Lefevre[1]

[1] Université de Lyon
Université Lyon 1, CNRS, LIRIS, UMR5205, F-69622, France
{Carole.Eyssautier, Stephanie.Jean-Daubias, Marie.Lefevre}@liris.cnrs.fr
[2] Université de Grenoble
LIG, BP 53, F-38041 Grenoble Cedex 9, France
Jean-Philippe.Pernin@imag.fr

Abstract. This paper considers the issues raised by the reuse of existing learners profiles. Indeed, heterogeneity of learners profiles, both structure and content wise, makes difficult their reuse in a system which haven't created them. A profiles modeling language (PMDL) is proposed in the aim of sharing and reusing profiles. Throughout this paper, the example of an existing profile is used to illustrate our propositions. A deployment of this language within a computerized environment have been carried out, which permitted to assess these propositions. Finally, evaluations are presented: one focused on PMDL language expressive power and the other on its deployment.

Keywords: human learning, learner models, learner follow-up, profiles modeling language.

1 Introduction

The national and international educational institutions are looking for common representations of data concerning learning activities. This is shown through numerous initiatives around skills frameworks, standardization works [1] or portfolio [2][3]. Some ILE (Intelligent Learning Environments) collect learning information in order to customize the system, help the teacher or tutor in his follow-up task or deliver a reflexive learning report to the learner. The teacher collects information to evaluate the learner, propose individualized activities or provide a view of the learning process to the learning situation actors. These various needs lead to different models of the individual information related to the learning. We call these models "learners profiles".

Currently, ILE designers and teachers build heterogeneous learners profiles according to their needs. There is no generic method allowing to reuse existing profiles – created by other systems, human or not – in different contexts or practices. Our research aims to fill this gap.

This paper first gives an overview on the issue of reusing heterogeneous profiles. Then the profiles modeling language PMDL is proposed. The deployment of this language within a computerized environment is presented as well as the results of our first evaluations. Finally, the implication of our contribution and the new perspectives opened by this work are discussed.

2 Heterogeneous Learners Profiles Reusing

In ILE research domain, the more general term to speak about personal information on a learner is "learner model". However, we prefer to use "learner profile" in the same sense than [4], [5] or [6]. We define a learner profile as *information concerning a learner or a group of learners, collected or deduced from one or several pedagogical activities, computerized or not. Information contained in the learner profile can concern his knowledge, abilities, conceptions or his behaviour* [7].

The heterogeneity of existing profiles – both structure and content – makes difficult their reuse in various contexts. The profiles reusing requires to harmonize their structure [8]: it consists in writing the various profiles according to a common formalism.

One way of solving the heterogeneity problem consists in defining a priori a set of information about a given learner. Normalization works have chosen this approach, like PAPI [9], IMS-RDCEO [10], or IMS-LIP [11]. However, these standards do not provide details necessary to describe information about learner's knowledge. Moreover, existing standards are focused on storage and exchange of data in order to help in managing educational institutions. This explains why standards are not precise enough to describe learner pofiles [4]. Finally, these information are stored as free text making difficult data reusing by a computer system [4]. This is a major limitation in our context of profiles reusing.

Another approach (as proposed in this paper) consists in reusing external profiles within a common computer framework by rewriting them according to an internal formalism. ViSMod [12] and DynMap systems [13] allow the reuse of learners profiles built by an ILE and based on an overlay model. They rewrite on one hand the domain-centric data, on the other hand the specific learner data. DynMap system [13] focuses on activity with the description of educational tools required by this activity and its specific evaluation criteria. The approach proposed by Zapata-Rivera *et al.* [12] requires first the description of the Bayesian network by the teacher.

None of these systems totally fulfills the need of reusing existing learners profiles. First, in order to encourage the integration of learners profiles in teacher's practices, it is necessary for him to be the major contributor in the whole profiles reuse process. This requires that he be able to associate profiles he has created with other ones built by other teachers or by ILEs. Moreover, such approach has to be built around knowledge and skills instead of being centered on activities: activities are specific to a

learning situation, not knowledge and skills. Finally, by reproducing domain specific data, previous systems do not allow to represent all information existing in learner profiles, such as free text (e.g. a comment), graphs with values suited to the links, or distribution lists [14].

3 Principles of PMDL Language

This section presents the principles of PMDL, Profiles MoDeling Language, we proposed in a recent PhD thesis [14]. This language allows to rewrite existing profiles in order to reuse them in a separate system. Section 3.1 presents the general structure of PMDL language. Section 3.2 shows an example of a profile rewritten according to PMDL specifications. Finally, section 3.3 proposes the specifications of one PMDL element, used in the previous example.

3.1 General Structure of PMDL

In order to propose PMDL language, a state of the art have first been carried out. Studied profiles come from research works, marketplace and teachers practices. For this study, we worked with seven teachers, coming from primary school to university and continuing education. The aim of this collaboration was to collect teachers and institutional practices about learners profiles. From this state of the art, we have analyzed in depth 24 profiles, selected for their richness and variety. We proposed to classify contained data in five main categories from which we based PMDL profiles modeling language. State of the art and classification are visible in [14].

Specifications of PMDL language are described in BNF (Backus Naur Form) formal notation. In this paper, in order to help understanding, a graphical notation[1] is also given, based on the notation used in some IMS specifications (like [11]). Figure 1 presents the general structure of PMDL according to the graphical representation.

1 In our case, graphical representation is less accurate than BNF notation, but easier for reading and understanding the concepts used.

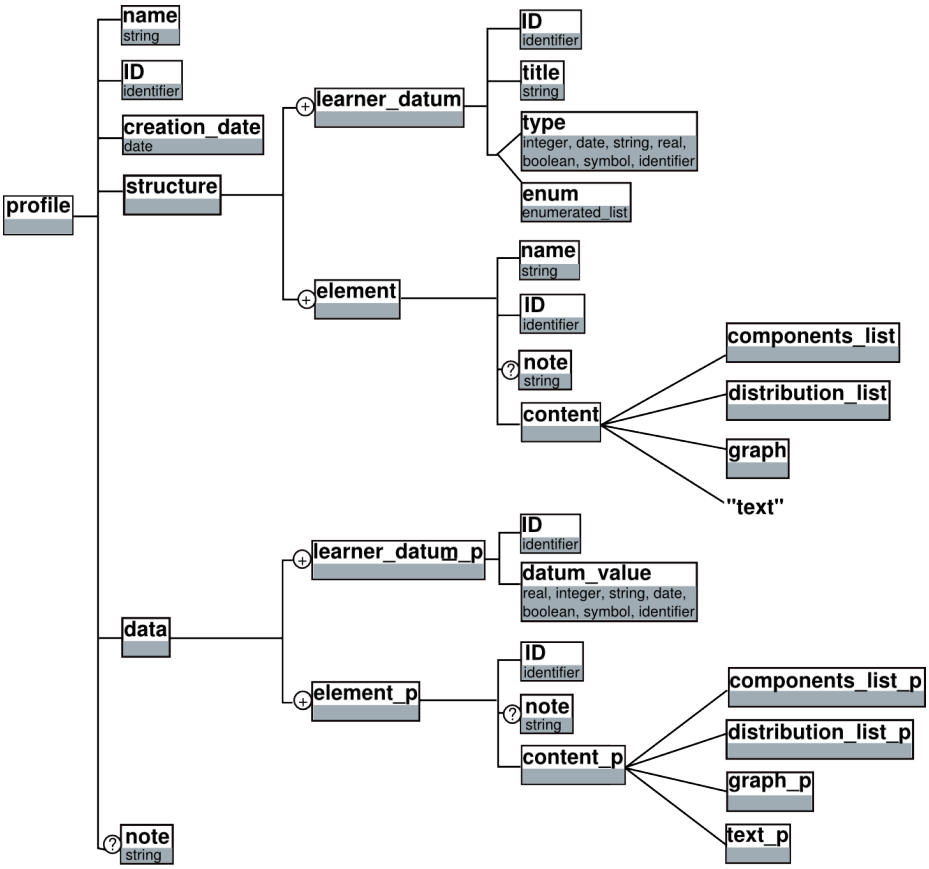

Fig. 1. General structure of PMDL

In PMDL, a *profile* consists of a *name*, an *ID*, a *creation-date*, a *structure* part, a *data* part and an optional *note*. In a learner profile written in PMDL, its *structure* and the learner *data* are described separately. Several learners profiles can then have the same *structure*, but different *data*.

***structure* part** consists of one or more general informations about a learner: *learner-datum*. A *learner-datum* contains an *ID* (e.g. D1), a *title* (e.g. "living city") and either a *type* (in the previous example "string") or an enumerate list, *enum* (e.g. "Lyon, London, Geneva"). s*tructure* part consists also of one or more *element*. An *element* is made of a *name* (e.g. "algebra"), an *ID*, an optional *note* and a *content*. The content can be of four types: *components_list*, *distribution_list*, *graph* and *text*. These, plus *learner_datum*, corresponds to the five categories identified in the state of the art. PMDL defined specifications for each one of these categories. In this paper, as an example, *distribution_list* specifications will be presented.

***data* part** consists of a set of data concerning a given learner, as well as one or more *element_p*. A *learner_datum_p* contains an *ID* (in the previous example: D1) and a *datum_value* (e.g. "London"). There must be as many *learner_datum_p*

(respectively *element_p*) as *learner_datum* (respectively *element*) previously declared. In addition, for each *ID* of a *learner_datum_p*, a *learner_datum* having the same *ID* must be previously declared. Identically, for each *ID* of an *element_p*, an *element* having the same *ID* must be declared in the *structure* part. The type of *content_p* must also be the same than the *content* type previously declared for this *ID*.

3.2 An Example of a Profile Rewritten in PMDL

To illustrate PMDL specifications, this section presents an example of an existing profile rewritten according to PMDL. Figure 2 presents a part of MoreMaths profile [15]. This profile represents the distribution of learner answers between correct, wrong and unanswered items for three components: "Introduction", "Polynomial division" and "Factorising polynomials". In this example, each component have been assessed by 10 questions. For the first component "Introduction", the learner made 4 correct answers, 6 wrong answers and 0 unanswered.

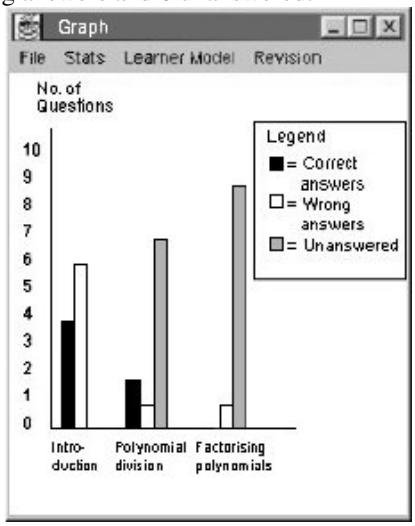

Fig. 2. A part of MoreMaths profile [15]

In order to rewrite this part of MoreMaths profile according to PMDL, *distribution_list* element have to be used. Indeed, it allows to represent the distribution of learner answers between some components. Figure 3 presents the *structure* part of MoreMaths profile rewritten according to PMDL and figure 4 presents its *data* part.

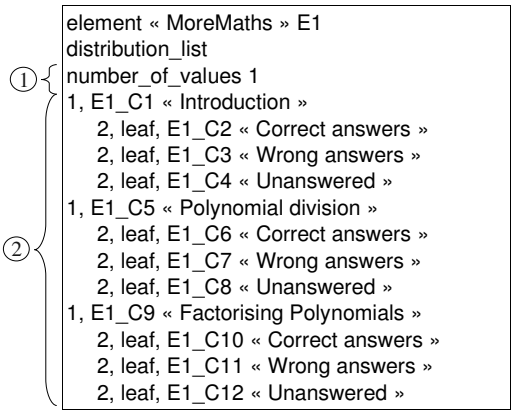

Fig. 3. Structure of MoreMaths profile in PMDL

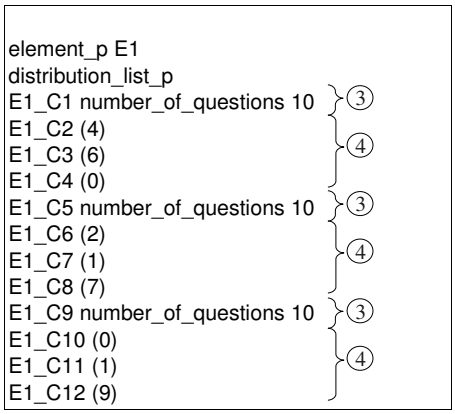

Fig. 4. Data of MoreMaths profile in PMDL

Figure 3 shows the ***structure*** of MoreMaths profile rewritten according to PMDL. The two first lines give the *element name* (MoreMaths) with its *ID* (E1) and its type of *content* (distribution_list). In the MoreMaths example (figure 2), there is only one value linked to each subcomponent ("correct answers", "wrong answers" and "unanswered"). In accordance with PMDL specifications, the tag "number_of_values" precedes an integer, 1 in example ①. Then, the structure of the components tree ② is rewritten according to PMDL specifications. At the *depth* 1, the component are described with an *ID* (e.g. E1_C1) and a *title* (e.g. "Introduction"). At the *depth* 2, subcomponents are described. *leaf* tags indicate leaf nodes in the tree. Following these tags, *ID* (e.g. E1_C4) and *title* (e.g. "Unanswered") can be found.

Figure 4 presents the ***data*** part of MoreMaths profile rewritten in PMDL. First line gives the *ID* of one element previously declared in the *structure* part of the profile (in this example the element E1, ie the MoreMaths element). Then, *ID* of each components declared in the structure part are listed. *ID* of components which have a depth 1 declared in the structure part (e.g. E1_C5) are followed by the tag "number_of_questions" and an integer. This indicates the number of questions in which the distribution is realised, 10 here ③. *ID* of components which have a depth 2 declared in the structure part (e.g. E1_C6) are followed by the value linked to the component in the profile ④. In the MoreMaths example, for subcomponents of "Introduction", value attributed to E1_C2 ("correct answers") is 4, E1_C3 ("Wrong answers") is 6 and E1_C4 ("Unanswered") is 0, and so on for the two other components. This *data* part gives the results obtained by the learner in example figure 2.

3.3 Specifications of the *distribution_list* Element

This section presents the PMDL specifications of the *distribution_list* element used in the previous example (section 3.2).

Structure Part of the *distribution_list* Element

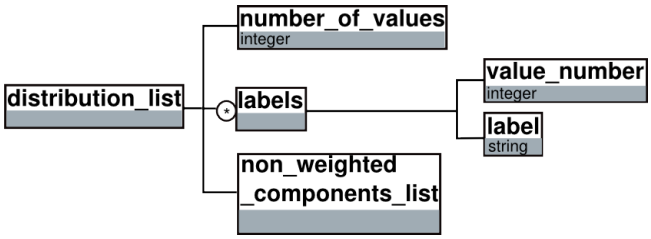

Fig. 5. Structure of *distribution_list* element

Figure 5 presents the graphical representation of *distribution_list structure* part. Due to a lack of space, we don't detail graphical representation of the last element, *non_weighted_components_list*, but we give its details in BNF below.

The first frame of BNF below gives the formal specifications corresponding to the figure 5. A tag "number_of_values" precedes an integer giving the number of values attributed to each components (① in figure 3). *sep* is a separator to distinguish the different elements. It can be for example instantiated in ";", or "crlf" (Carriage Return Line Feed). In the previous example, it was instantiated in crlf. An optional *labels* precedes *non_weighted_components_list*. There are no *labels* in MoreMaths example. *labels* is a set of pair <value_number> <label> where *value_number* is an integer and *label* consists of a tag "label" followed by a string. A label is used to give sense to the diverse values linked to a component:

1 label "distribution value"

2 label "success rate"

This example means that the first value linked to a component is the distribution value and the second one is the success rate of the learner to this component.

```
<distribution_list>
  ::= "number_of_values" <number_of_values> sep <non_weighted_components_list>
  ::= "number_of_values" <number_of_values> sep <labels> <non_weighted_components_list>

<number_of_values>  ::= <integer>

<labels>     ::= <value_number> <label> sep <labels>
             ::= <value_number> <label> sep
<value_number>  ::= <integer>
<label>      ::= "label" <string>
```

The frame below gives the BNF specifications for the element *non_weighted_components_list* which is a list of *non_weighted_component* (② in figure 3). This last one first gives the depth of the component in the tree of components[2]. If the component is a leaf of the tree of components, a tag "leaf" precedes the *component*. If it is not a leaf node of the tree, there is no tag and the

2 In order to help understanding, tabulations have been added in figure 3 to represent the tree of components. In the specifications, there are not.

component is followed by its child nodes, *non_weighted_subcomponents*. An optional *note* can be linked to each *component*. A component is defined by an *ID* (e.g. E1_C3) followed by a *title* (e.g. “Wrong answers”).

```
<non_weighted_components_list>
   ::= <non_weighted_component> sep <non_weighted_components_list>
   ::= <non_weighted_component> sep

<non_weighted_component>
   ::= <depth> "," "leaf," <component>
   ::= <depth> "," <component> sep <non_weighted_subcomponents>
   ::= <depth> "," "leaf," <component> "," <note>
   ::= <depth> "," <component> "," <note> sep <non_weighted_subcomponents>

<non_weighted_subcomponents>
   ::= <non_weighted_component> sep <non_weighted_subcomponents>
   ::= <non_weighted_component>

<depth>  ::= <integer>

<component>   ::= <ID> <title>
<ID>       ::= <identifier>
<title>     ::= <string>
```

Data Part of the *distribution_list* Element

The *structure* part of the *distribution_list* element have been presented previously. Its *data* part will be presented in this section. First, figure 6 shows the graphical representation of distribution list *data* part: *distribution_list_p*.

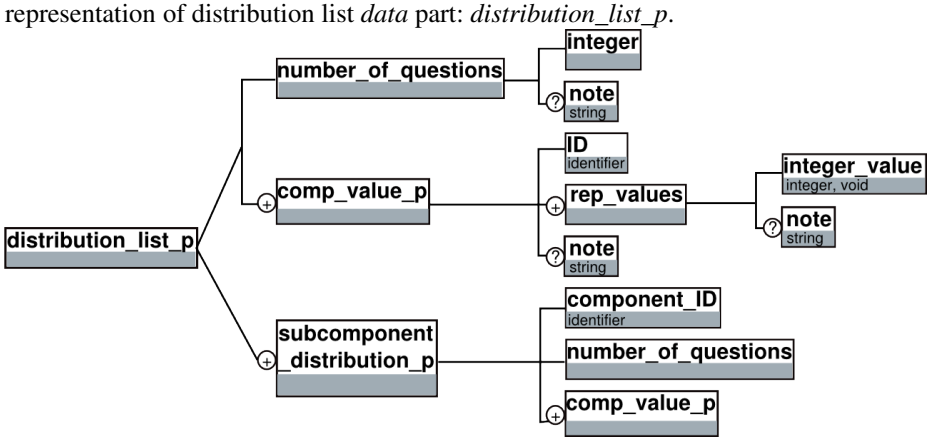

Fig. 6. Structure of *distribution_list_p* element

A *distribution_list_p* element may be of two types. The first one corresponds to the case of the values distribution on the components (depth 1), and the second one corresponds to the case of the distribution on the subcomponents (depth 2) (like the MoreMaths example figure 2). In the first case, the tag “number_of_questions” is followed by an integer in which the distribution is realised and by the list of components with their distribution values. In the second case the number of questions

(used to realised the distribution of the learner answers) is linked to the components which have been declared at a depth 1 in the *structure* part (③ in figure 4) and the values are associated to the components which have a depth 2 (④ in figure 4). *comp_value_p* is an *ID* of a *component* declared in the *structure* part followed by a list of values, *rep_values*. These values are void or integer (a distribution value is a count value so it is an integer). They can be linked to a *note*.

```
<distribution_list_p>
        ::= "number_of_questions" <number_of_questions> sep <component_distribution_p> //
case of a distribution on the components.
        ::= <subcomponent_distribution_p> // case of a distribution on the subcomponents

<number_of_questions>  ::= <integer>
                           ::= <integer> <note>

<component_distribution_p>  ::= <comp_value_p> sep <component_distribution_p>
                                ::= <comp_value_p> sep

<comp_value_p>  ::= <ID> "(" <rep_values> ")"
                    ::= <ID> "(" <rep_values> ")" <note>

<rep_values>       ::= <integer_value> ";" <rep_values>
                    ::= <integer_value>

<integer_value>   ::= <integer>
                    ::= <void>
                    ::= <void> <note>
                    ::= <integer> <note>

<subcomponent_distribution_p>
    ::=      <component_ID>      "number_of_questions"      <number_of_questions>      sep
<component_distribution_p> <subcomponent_distribution_p>
    ::=      <component_ID>      "number_of_questions"      <number_of_questions>      sep
<component_distribution_p>

<component_ID> ::= <identifier>
```

4 Implementation of PMDL

The aim of PERLEA project is to study the reuse of existing profiles by different actors of the learning situation [7]. EPROFILEA software (Exploitation of PROFILes by tEachers and leArners) has been developed within the framework of PERLEA project. Within this ILE, Bâtisseur module allows a teacher, or an educational designer, to describe the *structure* part of a profile he wants to reuse according to PMDL. Indeed, Bâtisseur interfaces are dynamically generated according to a DTD complying with PMDL specifications. *Data* part, containing the learner data, is filled in other modules of EPROFILEA.

Bâtisseur uses the metaphor of a brick wall construction to guide teacher in the description of a profiles structure. Four types of bricks are proposed to the teacher in Bâtisseur corresponding to the four types of *element content* in PMDL. Data of the last element described in PMDL (*learner_datum*) are managed globally in EPROFILEA. Indeed, such data have been considered to be related to the learner, not to one of his specific profiles. The implementation of the structure of a *distribution_list* element within Bâtisseur is presented figure 7, filled with the data of the MoreMaths example (figure 2).

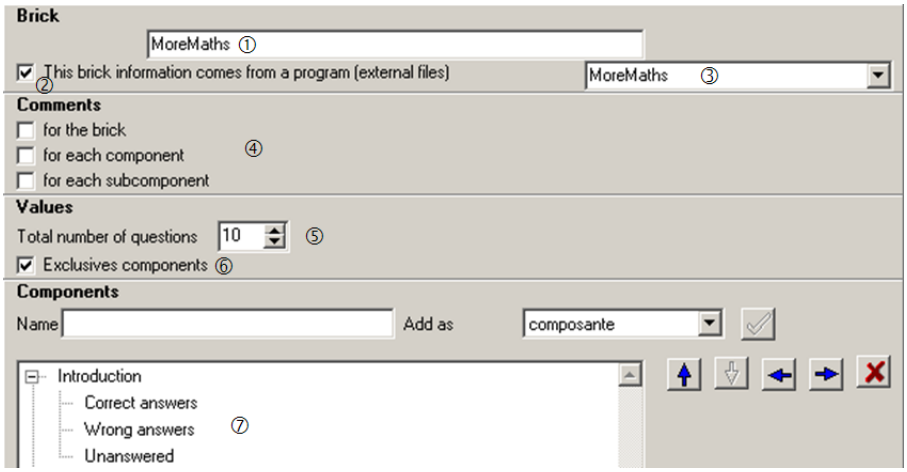

Fig. 7. A part of *distribution_list* element in Bâtisseur

First, teacher has to give the name of the *distribution_list* element ① (*name* field of *element* in PMDL). He indicates the origin of learner data ② (pencil and paper profile or built by an ILE) and if necessary adds from which software these data come from ③ (these two fields are specific to EPROFILEA and do not exist in PMDL specifications). Then he indicates if some comments ④ will be associated to the element and/or to each level of the components tree (correspond to *note* of *element* in PMDL and *note* of *component*). Teacher also gives the total number of questions ⑤ in which the distribution will be carried out (10 in the MoreMaths example). In Bâtisseur, teacher can specify if the distribution is exclusive or not ⑥: in the MoreMaths example, a learner cannot have both correct answer and wrong answer (do not exist in PMDL specifications). Finally, teacher can describe the components tree ⑦. He gives components names (*title* field of the *component*) and add their to the components tree as components or subcomponents (*depth* associated to the *component* in PMDL specifications).

Profiles structure created by a teacher with Bâtisseur is stored as an XML file. In EPROFILEA, integration of learners data in this file is either automatic when data come from an ILE – through the Tornade module – or manual, when data come from a pencil and paper profile – through the Prose module [7].

5 Evaluations of PMDL

Two types of PMDL evaluations are presented below: the first one focused on PMDL language expressive power and the other on its deployment within Bâtisseur module.

First, an application framework of PMDL language has been defined [14]. It qualifies which type of information can be described with PMDL: for example, raw data (as log files) cannot be described with PMDL, while analyzed data about learner knowledge can. Ten criteria have thus been defined. Then, PMDL language expressive power has been evaluated within this application framework, for the 24 studied profiles, as MoreMaths example presented in this paper.

All of these profiles consisted of at least one *learner_datum* element in order to identify the learner. *components_list* element is used in all studied profiles coming from teachers practices or from the marketplace. Indeed these profiles are based on skills frameworks which are lists of components. *text* element is used in teachers practices to add comments about a learner as well as in several profiles coming from research works (for example to describe misconceptions like in C-Polmile [15]). *graph* element is used in a few profiles, mostly to represent Bayesian networks. Finally, *distribution_list* element is seldom used (only 2 profiles among 24). These last two elements only exist in studied profiles coming from research works.

The only data existing in the studied profiles which cannot be described with PMDL are set outside of application framework. For example, profiles using Bayesian networks (like ViSMod [12]), cannot be rewrited entirely with PMDL: indeed, PMDL allows to described learner data about knowledge but not about the knowledge model which created them.

The implementation of PMDL within EPROFILEA has also been evaluated. Throughout this research, we have collaborated with seven teachers, coming from primary school to university and continuing education. Four of them, “partner teachers”, have been implicated in the design of several EPROFILEA modules and took part in their experimentation. Two of them were coming from primary school, one from junior high school and the last one from secondary school. They tested Bâtisseur to enter profiles coming from their practices. This experimentation took place in research laboratory with one interviewer, who was sometimes assisted by someone in charge of taking notes. The two primary school teachers were mostly met together in order to compare their point of views. First software tests meetings were devoted to presentation of EPROFILEA modules to get teachers feedback. Following meetings consisted in observing how teachers were using the software without giving them instructions. A fifth teacher, not involved in our work, also experimented the use of the module to enter her learners profiles.

First, the metaphor of a brick wall construction have been well accepted and understood by the teachers. It helped them to understand the aim of this module and to take it in hand quickly. Bricks more naturally used were Comments (*text* in PMDL) and List (*components_list* in PMDL) for which teachers found numerous examples in

their practices. In contrast, teachers had difficulties to understand and to use Graph (*graph* in PMDL) and Distribution (*distribution_list* in PMDL) bricks; examples were difficult to find in their practices. These results are consistent with conclusions of PMDL evaluation: we have not found such structure of data in studied teachers profiles. Teachers use instead skills frameworks, based on lists of components. Bâtisseur module was welcomed by teachers, in spite of the theoretical aspect of profiles structure. The environment is planed to be experimented in a classroom context with teachers not involved in our research. The question of using such a device by numerous teachers will be studied.

6 Conclusion

Many human actors and systems produce information about learning, but reusing these heterogeneous data is currently difficult. The plateform-independant Profiles Modeling Language (PMDL) we propose aims to solve this problem.

Based on a large state of the art, a classification of learners profiles data in five main categories has been used to define PMDL. This language allows to rewrite various learners profiles created by any systems, human or not. Existing profiles will now be able to be reused in transformations or exploitations purposes. In this article, an example with an existing learner profile illustrates the rewriting principle.

First results of PMDL evaluations are promising for language expressive power point of view. From an operational point of view, we have implemented this language within a profile management system named EPROFILEA. First experimentations seem to prove that teachers could use the main concepts of the language in order to represent, combine and exploit effective data, digital or not, constituting learner profiles.

The BNF formal notation used for the specifications of PMDL makes it usable in various contexts by different softwares. A potential use of this language is to format profiles as soon as they are created. Indeed, PMDL could be used as a first proposition for future works on profiles standardization.

References

[1] IMS, http://www.imsglobal.org/, (latest consultation 2008).

[2] Barrett, H.C.: Electronic Portfolios, In: Educational Technology : An Encyclopedia. ABC-CLIO (2001)

[3] Klenowski, V.: Developing portfolios for learning and assessment : processes and principles. RoutledgeFalmer (2002)

[4] Keenoy, K., de Freitas, S., Levene, M., Jones, A., Brasher, A., Waycott, J., Kaszas, P., Turcsanyi-Szabo, M., Montandon, L.: Personalised trails and learner profiling within e-Learning environments. European research network Kaleidoscope D22.4.1. (2004)

[5] Vassileva, J., McCalla, G., Greer, J.: Multi-Agent Multi-User Modeling. Special Issue of UMUAI, Vol. 13:(1), p. 179-210. (2003)

[6] Villanova-Oliver, M.: Adaptabilité dans les systèmes d'information sur le web : Modélisation et mise en œuvre de l'accès progressif. PhD of INPG, Grenoble, France. (2002)

[7] XX, An environment helping teachers to track students' competencies. In workshop LeMoRe, AIED'05, Amsterdam, Netherlands. (2005)
[8] XX, A model of learners profiles management process. In 14th International Conference on Artificial Intelligence in Education AIED'09, Brighton, UK. (2009) (To appear)
[9] PAPI, Draft standard PAPI learner. (2002)
[10] IMS RDCEO, Final Specification v1.0. (2002)
[11] IMS-LIP, Final Specification v1.0. (2001)
[12] Zapata-Rivera, J.D., Greer, J.: Interacting with Inspectable Bayesian Student Models. In IJAIED, Vol. 14, p. 1-37. (2004)
[13] Rueda, U., Larrañaga, M., Arruarte, A., Elorriaga, J.A.: DynMap+: A Concept Mapping Approach to Visualize Group Student Models, In EC-TEL'06, W. Nejdl & K. Tochtermann (Eds), p. 383-397, Crete, Greece. (2006)
[14] XX, Modèles, langage et outils pour la réutilisation de profils d'apprenants. PhD of Joseph Fourier University, Grenoble, France. (2008)
[15] Bull, S., McEvoy, A.T., Reid, E.: Learner Models to Promote Reflection in Combined Desktop PC / Mobile Intelligent Learning Environments. In Workshop LeMoRe, AIED'03, p. 199-208, Sydney. (2003)