
Uniformisation de la structure de profils d'apprenants issus de sources hétérogènes

Stéphanie Jean-Daubias^{*}, Carole Eyssautier-Bavay^{*,**},
Marie Lefevre^{*}

^{*} Université de Lyon, CNRS

Université Lyon 1, LIRIS, UMR5205, F-69622, France

^{**} Université Joseph Fourier, Grenoble

LIG, BP 53, F-38041 Grenoble Cedex 9, France

{Stephanie.Jean-Daubias, Carole.Eyssautier, Marie.Lefevre}@liris.univ-lyon1.fr

RÉSUMÉ. Nous nous intéressons à la réutilisation par des acteurs autres que leurs créateurs, de profils d'apprenants issus d'EIAH ou de pratiques d'enseignants, dont nous ne connaissons ni le contenu ni la structure. Dans cet article, nous présentons les enjeux de cette question et notre approche pour y répondre. Nous exposons les travaux traitant ces enjeux, la solution que nous proposons, ainsi que sa mise en œuvre. Nous présentons en effet PMDL, le langage de description de profils que nous proposons pour rendre possible la réutilisation de profils d'apprenants externes, papier-crayon ou issus d'EIAH. Nous présentons également la mise en œuvre de ce langage dans un environnement permettant notamment à des enseignants de décrire des structures de profils. Nous illustrons cet exposé en prenant l'exemple de la réécriture du profil de l'EIAH MoreMaths avec le formalisme que nous proposons. Nous terminons sur les évaluations qui ont été conduites, d'une part concernant l'expressivité du langage proposé, d'autre part concernant l'opérationnalisation de ce langage.

MOTS-CLÉS : profils d'apprenant, langage de description de profils, assistance à l'enseignant.

1. Contexte et travaux connexes

Cette recherche s'intéresse à la réutilisation par des acteurs autres que leurs créateurs, de profils d'apprenants issus d'EIAH ou de pratiques d'enseignants, dont nous ne connaissons ni le contenu ni la structure. Nous définissons un profil d'apprenant comme un ensemble d'informations caractérisant ses connaissances, compétences, conceptions, ou son comportement. La réutilisation de profils nécessite l'harmonisation de leurs structures qui consiste à écrire les divers profils dans un formalisme commun, rendant possible des traitements variés, mais communs, sur ces profils. La difficulté et l'intérêt de cette problématique résident dans la recherche d'une approche *générique* (prenant en compte la diversité des profils), *souple* (s'adaptant aux formalismes existants), et *adaptée aux utilisateurs*, principalement des enseignants. Notre approche répond à ces contraintes : nous avons en effet établi un langage générique de description de profils d'apprenants permettant de décrire la plupart des profils existants, quels que soient la discipline ou le niveau concernés. Nous avons opérationnalisé ce langage au sein d'un environnement adapté aux enseignants, leur permettant de convertir les profils externes qu'ils utilisent pour les rendre conformes au formalisme utilisé dans notre environnement, ce qui permet ensuite de proposer des exploitations mutualisées de ces profils. Dans cet article, après les enjeux de cette question et notre approche pour y répondre, nous présentons les travaux traitant ces enjeux, la solution que nous proposons, sa mise en œuvre et leur évaluation.

Une des solutions au problème de la réutilisation de profils consiste à définir de façon consensuelle *a priori* ce qu'est un ensemble d'informations sur l'apprentissage d'un individu. C'est l'approche adoptée par les travaux sur la normalisation des données personnelles des apprenants [PAPI 02] [IMS_RDCEO 02] [IMS-LIP 01]. Elle favorise la réutilisation de profils, mais uniquement pour ceux qui respectent la norme concernée, manquant ainsi de souplesse : en l'absence de norme universelle cette approche ne permet pas la réutilisation de tous les types de profils. Par ailleurs, les informations les plus pertinentes dans notre contexte ne sont pas décrites précisément dans les standards existants. En effet, un profil représente les informations sur les connaissances de l'apprenant à un niveau de granularité fin, alors que les standards visent à faciliter l'échange des données, ce qui implique un niveau de granularité plus élevé. De plus, ces informations, stockées sous forme de texte libre, ne sont pas facilement exploitables par un système informatique [KEENOY et al. 04]. Une seconde approche, dans laquelle nous nous inscrivons, vise à réutiliser des profils externes dans un unique environnement informatique en les réécrivant *a posteriori* selon un formalisme interne. Ainsi, les systèmes ViSMod [ZAPATA-RIVERA & GREER 04] et DynMap [RUEDA et al. 06] réécrivent d'une part les données du domaine étudié, et d'autre part les données de l'apprenant considéré, mais manquent de généralité puisqu'ils ne permettent pas de représenter les informations sous forme de texte en langage naturel, de graphes avec valeurs portées par les arcs, de composantes à valeurs issues de répartition. Enfin, ces systèmes laissent peu de place aux enseignants et ne prennent en compte que les profils issus d'EIAH.

2. Principes du langage PMDL

Notre approche de la réutilisation de profils externes hétérogènes passe par une harmonisation qui conduit à une représentation unifiée de la structure et des données, ce qui nécessite un formalisme interne commun de description des profils. Après une revue des profils existants provenant d'EIAH issus de la recherche, du marché et de pratiques d'enseignants, nous avons sélectionné vingt-cinq profils représentatifs, étudiés de façon approfondie pour proposer une catégorisation des informations qu'ils contenaient. Nous avons ainsi répertorié cinq catégories d'informations nous permettant d'établir un langage de description de profils : PMDL (Profiles MoDeling Language) [EYSSAUTIER-BAVAY 08].

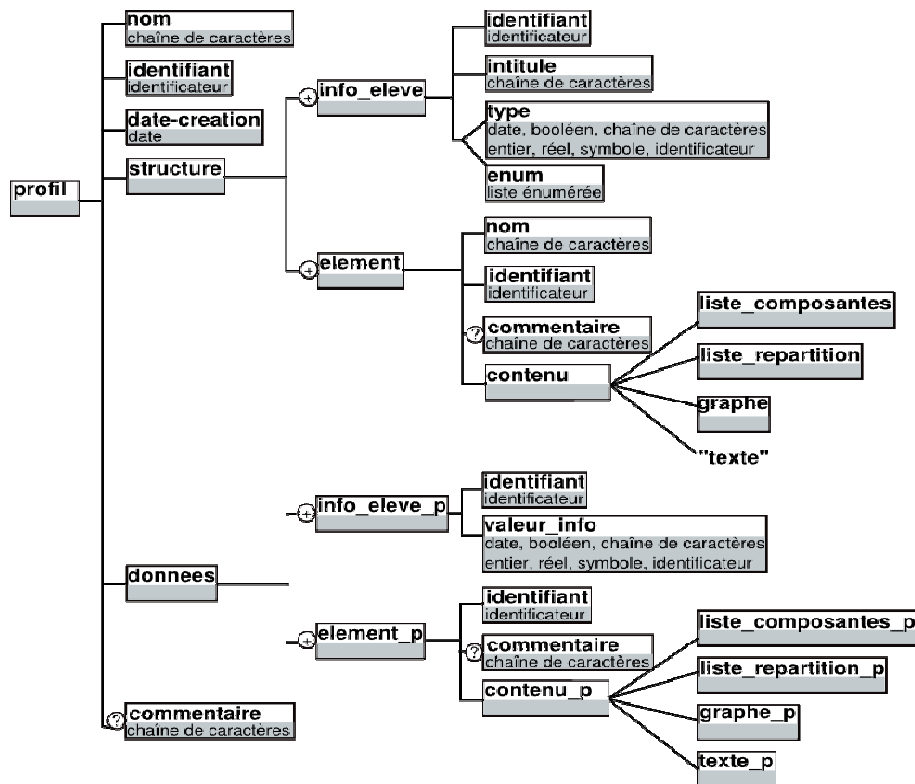
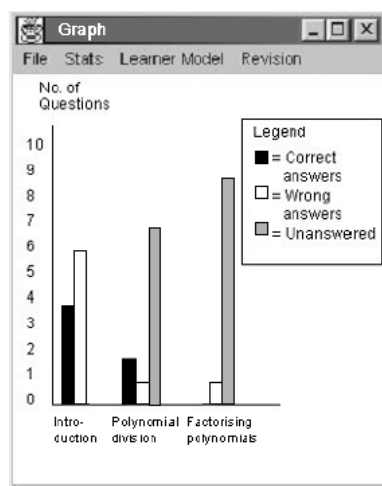


Figure 1. Structuration générale simplifiée de PMDL

La figure 1 présente la structuration générale du langage PMDL, qui décrit un *profil* comme étant constitué d'un *nom*, d'un *identifiant*, d'une *date-creation*, d'une partie *structure*, d'une partie *donnees* et éventuellement d'un *commentaire*. La **partie structure** est composée d'une ou plusieurs informations générales sur l'élève, *info_eleve*. Une *info_eleve* est constituée d'un *identifiant*, d'un *intitule* de type chaîne de caractères (par exemple « ville de résidence ») et soit d'un *type* (« chaîne de caractère » dans notre exemple), soit d'une liste énumérée, *enum* (par exemple

« Lyon, Villeurbanne, Bron »). La partie *structure* des profils est également composée d'un ensemble non vide d'*element*. Un *element* est constitué d'un *nom* (par exemple « algèbre »), d'un *identifiant*, éventuellement d'un *commentaire*, et d'un *contenu* qui peut être de quatre types : *liste_composantes*, *liste_repartition*, *graphe* et *texte*. Ces quatre types auxquels s'ajoute *info_eleve* correspondent aux cinq catégories d'informations identifiées dans notre étude de l'existant. Chacune de ces catégories donne lieu à des spécifications au sein du langage PMDL. Quant à la **partie données** du *profil*, elle est composée, comme la partie *structure*, d'un ensemble d'informations sur l'élève, ainsi que d'un ensemble d'*element_p*. Une *info_eleve_p* est constituée d'un *identifiant* et d'une *valeur_info*. Il existera autant d'*info_eleve_p* (respectivement *element_p*) que d'*info_eleve* (respectivement *element*) précédemment déclarés. Par ailleurs, pour chaque *identifiant* d'*element* et d'*info_eleve* déclaré, il existera un *element_p* reprenant cet *identifiant* et lui associant un *contenu_p* de même type que le *contenu* précédemment déclaré.



```

element « MoreMaths » E1
liste_repartition
nbre_valeurs 1
1, E1_C1 « Introduction »
  2, feuille, E1_C2 « Correct answers »
  2, feuille, E1_C3 « Wrong answers »
  2, feuille, E1_C4 « Unanswered »
1, E1_C5 « Polynomial division »
  2, feuille, E1_C6 « Correct answers »
...
element_p E1
liste_repartition_p
E1_C1 nbre_question 10
E1_C2 (4)
E1_C3 (6)
E1_C4 (0)
E1_C5 nbre_question 10
E1_C6 (2)
...

```

} S
T
R
U
C
T
U
R
E

} D
O
N
N
É
E
S

Figure 2. MoreMaths version mobile **Figure 3.** Profil MoreMaths en PMDL

Pour illustrer ce langage, prenons l'exemple de la version mobile du profil MoreMaths [BULL et al. 03] dont les informations sont de type *liste_repartition* (cf. figure 2). Ce profil donne la répartition des réponses de l'apprenant entre réponses correctes, réponses fausses et non réponses : les trois sous-composantes des composantes « Introduction », « Polynomial division » et « Factorising polynomials ». Ici, l'apprenant a 6 réponses fausses et 4 correctes pour « Introduction ». La figure 3 présente un extrait de ce même profil réécrit selon les spécifications du langage PMDL (spécifié en BNF afin d'être indépendant de toute technologie). Nous y trouvons la description de la structure de ce profil, puis la description des données de l'apprenant concerné. Dans la partie *structure*, les composantes de la *liste_repartition* sont décrites telles que définies dans MoreMaths : au niveau 1, on trouve les composantes « Introduction », « Polynomial division » et au niveau 2, comme feuilles de l'arbre des composantes, les trois sous-

composantes de chacune des composantes. Un identifiant est associé à chaque composante et sous-composante (E1_C1 pour la première composante). Dans la partie *données*, le nombre de questions concernées par la répartition des réponses de l'apprenant est associé aux composantes (10 dans l'exemple). Enfin, les valeurs sont associées aux composantes de niveau 2. On retrouve ainsi les 6 réponses fausses et les 4 réponses correctes de l'apprenant de l'exemple pour la composante « Introduction ». PMDL permet ainsi de réécrire des profils externes selon un formalisme commun. Nous montrons dans la partie suivante comment nous avons opérationnalisé ce langage.

3. Mise en œuvre de PMDL au sein de Bâtisseur

L'environnement EPROFILEA [JEAN-DAUBIAS & EYSSAUTIER-BAVAY 05] unifie les profils externes pour en faire des exploitations mutualisées. Pour cela le module Bâtisseur permet à un enseignant de décrire la partie *structure* du profil telle que spécifiée dans PMDL. La partie *donnees* contenant les informations des apprenants est renseignée par la suite dans des modules d'intégration de données.

Bâtisseur donne les moyens à l'enseignant de construire la structure des profils qu'il souhaite manipuler par interprétation des informations contenues dans les profils, papier-crayon ou logiciels, sur lesquels il s'appuie éventuellement. Bâtisseur repose sur la métaphore de la construction d'un mur de briques (la structure de profils). Quatre types de briques sont proposés : Commentaire, Liste, Répartition et Graphe. Ces types de briques respectent les spécifications des quatre principaux types d'*element* du langage PMDL. Les informations correspondant au cinquième type, *info_eleve*, sont gérées de manière globale au sein d'EPROFILEA. Les briques de type **Commentaire** (*texte* dans le langage de description de profils) permettent de représenter des informations sous forme textuelle. Elles correspondent à des zones de texte libre, sans aucune contrainte sur l'organisation des données qui y seront portées. Les briques de type **Liste** permettent de décrire des informations sous forme de liste hiérarchique (*liste_composantes* dans PMDL). Avec ce type de briques, un enseignant peut décrire ses informations sous forme d'un ensemble de *composantes*, chacune pouvant être détaillée en *sous-composantes*. Dans le profil de l'apprenant (partie *donnees*), les *valeurs* seront attribuées aux composantes de niveau hiérarchique le plus bas. Ces valeurs sont caractérisées par les informations définies par l'enseignant lors de la création de la brique (partie *structure*) : le nombre de valeurs associées à chacune des composantes (par exemple deux valeurs), leurs *echelles* (par exemple une note entre 0 et 20 et un nombre entier positif) et éventuellement l'*unite* correspondante (par exemple rien pour la première valeur et une durée de travail pour la seconde valeur). Les briques de type **Répartition** (*liste_repartition* du langage) permettent de décrire la répartition des réponses d'un apprenant entre différents comportements listés. De la même manière que dans les briques de type Liste, les comportements sont décrits à l'aide de *composantes*, qui peuvent être détaillées en *sous-composantes*. Ce type de briques permet de représenter le comportement d'un apprenant par rapport à des comportements attendus, en comptant le nombre de fois où celui-ci a mis en œuvre les différents

comportements. Dans ce type de briques, les composantes peuvent être exclusives. Dans ce cas l'enseignant doit préciser le nombre de questions concernées par la brique, ce qui permettra de contrôler les saisies de données. Les briques de type **Grphe** (*graphe* dans PMDL) permettent de décrire des informations sur l'apprenant sous forme de *composantes* (les sommets du graphe), et de *liens* qui peuvent exister entre elles (les arcs du graphe). Ces composantes ne sont pas décomposables en sous-composantes comme dans les briques des deux types précédents. Dans ce type de briques, on peut associer aux composantes et/ou aux liens entre composantes un certain nombre de *valeurs*, associées à des *echelle* et éventuellement complétées d'*unite*.

Figure 4. Brique "MoreMaths" de type Répartition dans Bâtisseur

Reprenons l'exemple de la brique MoreMaths : la figure 4 présente une partie de l'écran proposé à l'enseignant pour construire une brique de type Répartition. Notons que l'interface de Bâtisseur est générée dynamiquement en fonction d'une description opérationnelle de PMDL, qui prend la forme d'une DTD. Pour créer la brique « MoreMaths », de type de répartition, l'enseignant donne son nom ① (le *nom* de l'*element* dans PMDL), puis indique si les données des apprenants seront tirées de données papier-crayon ou de données informatisées ② (information non présente dans PMDL) et précise de quel logiciel sont issues les données ③. Ensuite, il indique si des commentaires ④ seront associés à chaque niveau de l'arbre des composantes et/ou globalement pour la brique (champs *commentaire* de l'*element* et de la *composante*). Il doit préciser le nombre total de questions ⑤ qui pourront être associées à l'ensemble des feuilles de l'arbre des composantes – chaque composante ou sous-composante de niveau le plus bas – (champ *nbre_valeurs* de l'*element liste_repartition*). Ici la répartition est exclusive ⑥ : un apprenant ne pourra pas avoir une réponse correcte en même temps qu'une réponse fautive (notion absente de PMDL). Une fois ces informations saisies, l'enseignant peut créer l'arbre des composantes ⑦ (expression du *niveau* de chaque *composante*) en précisant pour chacune son nom (champ *intitule* des *composante*).

```

<structure id="68" nom="Maths" createur="Marie" date_creation="17/07/2008"
date_derniere_modif="17/07/2008" nom_eleve="" prenom_eleve="" >
- <brique id="1" type="1" nom="MoreMaths" logiciel_externe="MoreMaths" date_evaluation=""
evaluation_source="" nbquestions="10" indice="0">
- <arbre_des_composantes niveaux="2" cumul="faux">
- <composante nom="Introduction">
- <sous_composante nom="Correct answers"><valeur>❶</valeur> </sous_composante>
- <sous_composante nom="Wrong answers"><valeur>❶</valeur> </sous_composante>
- <sous_composante nom="Unanswered"><valeur>❶</valeur> </sous_composante>
</composante>

```

Figure 5. Extrait de la structure de profils de la brique "MoreMaths"

La figure 5 montre le résultat de la création de cette structure de profils, stocké sous forme d'un fichier XML par Bâtitseur. Cette structure de profils sera ensuite instanciée en complétant les valeurs ❶ avec les données de chaque apprenant dans un module d'intégration de données de l'environnement EPROFILEA.

4. Bilan et évaluations

Dans cet article nous avons présenté PMDL, le langage de description de profils que nous proposons pour rendre possible la réutilisation de profils d'apprenants, papier-crayon ou issus d'EIAH ainsi que sa mise en œuvre. Ce travail a fait l'objet d'évaluations, d'une part concernant l'expressivité du langage proposé, d'autre part concernant l'opérationnalisation de ce langage au sein du module Bâtitseur.

Nous avons défini le cadre d'application de PMDL, puis évalué son expressivité dans ce cadre pour les 25 profils de notre étude préalable, en les exprimant avec notre langage. *Informations_eleve* est présent dans tous les profils, car tous comportent une identification de l'élève. *Liste_composantes* est utilisé dans tous les profils étudiés issus du marché et des pratiques, qui s'appuient sur des référentiels de compétences sous cette forme. Le *texte* libre est utilisé dans les pratiques des enseignants pour les commentaires et dans certains profils issus de la recherche pour exprimer des conceptions erronées. *Graphe* n'est utilisé que dans des profils issus de la recherche, la plupart basés sur des réseaux bayésiens. Enfin, *liste_repartition* est très rarement utilisé. Ces deux derniers *element* étant plus complexes et absents des référentiels de compétences, on ne les retrouve ni dans les profils du marché, ni dans les pratiques. Les seules informations présentes dans les profils étudiés et ne pouvant pas être représentées par PMDL se situent en dehors du cadre d'application défini. Les profils sous forme de réseaux bayésiens ne peuvent par exemple être représentés en intégralité par le langage : en effet, nous nous intéressons uniquement aux données de l'apprenant, indépendamment du modèle de connaissances qui en est à l'origine. PMDL est donc en mesure de représenter les données de l'apprenant à un instant t, mais pas les données liées à la modélisation des connaissances du domaine.

L'opérationnalisation du langage a également été testée : Bâtitseur permet de construire les structures correspondant à des profils existants, aussi bien papier-crayon que logiciels. En complément d'évaluations ergonomiques et de tests auprès d'utilisateurs hors public cible, Bâtitseur a été mis à l'essai en laboratoire auprès

d'enseignants ayant participé à la conception, puis testé auprès d'une enseignante extérieure au projet. Tout d'abord, la métaphore de la construction du mur de briques a été bien acceptée et bien comprise par les enseignants. Elle leur a permis de comprendre l'objectif de ce module et a favorisé une prise en main relativement rapide de l'outil. Les briques les plus naturellement utilisées ont été les briques *Commentaires* et *Liste* pour lesquelles les enseignants ont trouvé de nombreux exemples. Les briques *Grappe* et *Répartition* ont été plus difficilement assimilées, les exemples ayant été difficiles à trouver par les enseignants. Ceci est cohérent avec les conclusions de l'évaluation de PMDL : les référentiels de compétences utilisés par les enseignants ne comportant pas d'informations structurées ainsi. Le module de description de profils a ainsi été bien accueilli par les enseignants auprès desquels il a été testé, ce malgré l'abstraction de la notion de structure de profils manipulée. Des évaluations plus poussées, faisant appel à un plus grand nombre d'enseignants, dans un contexte s'approchant progressivement de conditions d'utilisation réelles, pourront être menées lorsqu'un prototype complet de l'environnement EPROFILEA sera disponible. La description de profils n'est en effet pas une fin en soit : elle ne trouve tout son intérêt que par les utilisations mutualisées des profils ainsi reformulés qu'elle rendra possible en aval [JEAN-DAUBIAS & EYSSAUTIER-BAVAY 05].

5. Bibliographie

- [BULL et al. 03] Bull, S., McEvoy, A. T., Reid, E., « Learner Models to Promote Reflection in Combined Desktop PC / Mobile Intelligent Learning Environments », *Workshop LEMORE, AIED'03*, Sydney, juillet 2003, p. 199-208.
- [EYSSAUTIER-BAVAY 08] Eyssautier-Bavay, C., Modèles, langage et outils pour la réutilisation de profils d'apprenants, Thèse de doctorat, Université Joseph Fourier, 2008.
- [JEAN-DAUBIAS & EYSSAUTIER-BAVAY 05] Jean-Daubias, S., Eyssautier-Bavay, C., « An environment helping teachers to track students' competencies », *Workshop LEMORE, AIED'05*, Amsterdam, juillet 2005, p. 19-23.
- [KEENOY et al. 04] 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, *Kaleidoscope D22.4.1*, 2004.
- [RUEDA et al. 06] Rueda, U., Larrañaga, M., Arruarte, A., Elorriaga, J.A., « DynMap+: A Concept Mapping Approach to Visualize Group Student Models », *Innovative Approaches for Learning and Knowledge Sharing, EC-TEL'06*, Crete, 2006, p. 383-397.
- [ZAPATA-RIVERA & GREER 04] Zapata-Rivera, J. D., Greer, J., « Interacting with Inspectable Bayesian Student Models », *IJAIED*, vol. 14; 2004, p. 1-37.

6. Références sur le WEB

- [IMS-LIP 01] Spécification finale v1.0 d'IMS-LIP, mars 2001, <http://www.imslobal.org/profiles/lipbest01.html>.
- [IMS_RDCEO 02] Spécification finale v1.0 d'IMS RDCEO, octobre 2002, <http://www.imslobal.org/competencies/>.
- [PAPI 02] Version de travail de la spécification PAPI, février 2002, <http://edutool.com/papi/>.