

Intégration de données hétérogènes : un exemple pour la constitution de profils d'apprenants

Marie Lefevre, Stéphanie Jean-Daubias

Université de Lyon, CNRS,
Université Lyon 1, LIRIS, UMR5205, F-69622, France
{prenom.nom}@liris.cnrs.fr

Résumé : Cet article aborde la question de l'intégration de données hétérogènes dont l'organisation n'est pas préalablement connue, dans des structures de données respectant un formalisme unique connu. Le contexte de ce travail est une recherche sur des profils d'apprenants dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain). Les structures de données traitées sont les profils issus de logiciels pédagogiques externes. Le principe du processus d'intégration que nous proposons s'articule en deux phases : une phase d'initialisation qui permet la prise en charge d'un nouveau logiciel pédagogique en constituant un convertisseur et une phase d'exécution du convertisseur créé sur les données des profils à convertir. Outre le processus d'intégration que nous avons défini, nous présentons dans cet article la mise en œuvre que nous en avons faites et son évaluation.

Mots-clés : intégration de données, convertisseur, profils d'apprenants, Environnements Informatiques pour l'Apprentissage Humain.

1. Introduction

L'intégration de données hétérogènes est un problème informatique que l'on doit aborder dès lors que l'on cherche à réutiliser des données non normalisées issues de sources différentes.

C'est notre cas dans le cadre d'un projet de recherche qui a pour but la proposition de modèles et d'outils favorisant la réutilisation de profils d'apprenants hétérogènes par des acteurs autres que leurs créateurs (Jean-Daubias et al. 2009).

Dans cet article centré sur l'intégration de données hétérogènes, nous présentons tout d'abord le contexte de notre recherche et la problématique identifiée. Après un aperçu des approches adoptées pour l'intégration de données, nous exposons le principe que nous avons établi. Nous montrons ensuite le système que nous avons développé et indiquons comment nous avons évalué ce travail.

2. Contexte

Ce travail s'inscrit dans le projet PERLEA, projet de recherche en informatique dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain) qui s'intéresse à l'ingénierie des profils d'apprenants.

Nous définissons un *profil d'apprenant* comme un ensemble d'informations interprétées, concernant l'apprentissage d'un apprenant ou d'un groupe d'apprenants, collectées ou déduites à l'issue d'une ou plusieurs activités pédagogiques, qu'elles soient ou non informatisées (Jean-Daubias *et al.* 2009). Notons que les données d'un profil sont définies selon une structure précise. Cette structure est indépendante des données d'un apprenant particulier et peut être partagée : elle peut être utilisée pour les profils de plusieurs apprenants, alors que les données sont personnelles et relèvent de l'apprenant ou du groupe d'apprenants concerné par le profil.

Le projet PERLEA vise à proposer des modèles et des outils permettant la réutilisation de profils d'apprenants issus de sources diverses. Pour cela, nous proposons un environnement, EPROFILEA, qui réorganise des profils d'apprenants externes hétérogènes avant de proposer des exploitations mutualisées des profils ainsi réorganisés¹ : visualisation de profils et personnalisation de l'apprentissage selon le contenu des profils (Jean-Daubias *et al.* 2009). L'idée est de proposer aux enseignants des outils facilitant la gestion et l'exploitation des différents profils auxquels ils ont accès.

Dans un but de réutilisation de profils d'apprenants existants, nous avons défini un langage de description de profils, PMDLe. Le langage PMDLe (evolutive Profiles MoDEling Language) est un langage de modélisation de profils qui permet de décrire la structure d'un profil afin d'exprimer différents profils hétérogènes selon un même formalisme (Jean-Daubias *et al.* 2009) (Jean-Daubias & Ginon 2010). Pour cela, PMDLe distingue dans un profil d'apprenant sa partie *structure*, pouvant être commune à plusieurs profils, de sa partie *données*, contenant les informations propres à chaque apprenant. Le langage comporte quatre éléments de base génériques, instanciables et combinables pour définir le modèle des profils à gérer.

Dans notre contexte, ce langage sert de langage pivot dans la représentation des profils. Ceci nous permet d'envisager l'intégration de données issues de profils externes, ne respectant pas ce langage. Deux cas sont possibles dans notre contexte : l'intégration de données papier (pour les profils informels qui peuvent être non écrits ou les profils établis sous forme papier) et l'intégration de données issues de logiciels (principalement pour les profils établis par des EIAH).

Dans les deux cas, l'intégration des données des profils externes est le lieu de la transformation du modèle de profils externes en un modèle conforme au langage de description de profils. Il s'agit donc d'aligner le modèle de profils source (en partie implicite dans le cas de profils non écrits, exprimé par

¹ Nous ne parlons pas de profils normalisés afin qu'il n'y ait pas de confusion possible avec la notion de normes et standards pédagogiques.

l'organisation des données du profil dans les autres cas), sur le modèle de profils destination (respectant PMDLe).

Pour l'intégration de profils logiciels, l'alignement des modèles se fait explicitement. Il nécessite la connaissance, non seulement du modèle de profils destination (structure de profils respectant PMDLe), mais également du modèle de profils source (les règles régissant l'organisation des données du profil externe à intégrer) pour chaque logiciel dont seront issus des profils à traiter.

Pour l'intégration des profils papier-crayon, cet alignement se fait au contraire implicitement lors de la saisie des profils dans une structure de profils conforme à PMDLe. Dans ce cas, c'est l'interface du module de saisie des profils qui est le lieu de l'alignement des modèles, et c'est l'utilisateur enseignant qui est l'acteur de cet alignement.

Dans la suite de cet article, après un état de l'art concernant l'intégration de données, nous présentons notre approche pour l'intégration de profils externes issus de logiciels.

3. Les approches de l'intégration de données

L'intégration de données issues de sources hétérogènes est un problème informatique abordé dans de multiples domaines : recherche d'information, aide à la décision, web sémantique, et de manière plus générale, gestion des connaissances. L'hétérogénéité des données est due aux différents formats et structures de stockage. Ainsi, les données peuvent être issues de sources structurées comme des bases de données relationnelles, de sources semi-structurées comme des documents XML ou de sources non structurées comme des documents texte.

L'approche des *systèmes médiateurs*, consiste à définir une interface entre l'agent (humain ou logiciel) effectuant une requête et l'ensemble des sources de données accessibles (Wiederhold 1992). Le médiateur comporte un schéma global (modèle du domaine d'application et vocabulaire structuré pour l'expression des requêtes), des vues abstraites (décrivant le contenu des différentes sources de données à l'aide du vocabulaire structuré) et des adaptateurs. Lorsqu'une requête est posée au médiateur, celui-ci la traduit en termes de vues et se sert des adaptateurs pour traduire ces vues dans le langage de requêtes accepté par chaque source de données. Un tel système donne l'impression d'interroger un système centralisé contenant des sources homogènes, alors que les sources sont réparties et hétérogènes. La correspondance entre le schéma global et les schémas des sources de données à intégrer peut être faite de deux manières (Hacid & Reynaud 2004) : en définissant le schéma global en fonction des schémas des sources de données à traiter, ou au contraire en définissant les schémas des sources en fonction du schéma global.

Dans notre contexte, nous disposons de l'équivalent d'un schéma global par l'intermédiaire du langage PMDLe et les sources de données existent déjà puisque nous voulons intégrer des profils d'apprenants issus de logiciels existants. Nous ne pouvons pas définir le schéma global en fonction des schémas des sources de

données, puisque celles-ci ne sont pas connues à l'avance. Nous ne pouvons pas non plus agir sur les schémas des sources de données externes sur lesquelles nous n'avons pas d'emprise. Nous ne pouvons donc pas retenir cette approche, mais conservons l'idée qu'il est nécessaire de définir un adaptateur entre le schéma global et chaque source de données.

L'approche des *entrepôts de données* consiste à intégrer et stocker dans un environnement les sources issues de systèmes distribués (Hacid & Reynaud 2004). Pour cela, un adaptateur est défini pour chaque source de données de manière à en extraire les données et à les transformer pour qu'elles soient compatibles avec le format de l'entrepôt. Des requêtes sur ces données peuvent ensuite être faites via l'entrepôt de données.

Un environnement de gestion de profils comme le nôtre peut être considéré comme un entrepôt de données, toutefois, nous ne souhaitons pas extraire, convertir et stocker toutes les données contenues dans les profils sources, mais seulement traiter les données nécessaires au moment opportun pour remplir les structures de profils utilisées par les enseignants, ce qui peut se faire en plusieurs fois : soit en fonction des besoins de l'enseignant, soit au fur et à mesure de la création ou mise à jour des profils externes. Un entrepôt de données n'est ainsi pas pleinement adapté à notre situation.

Dans le domaine des EIAH, des techniques de transformation de modèles ont été proposées pour la spécification de scénarios pédagogiques. Laforcade propose par exemple une traduction entre CPM, le langage qu'il a créé (s'appuyant sur des modèles UML), et des modèles IMS-LD (en XML) (Laforcade 2005). Cette transformation se réduit conceptuellement à une transformation de modèles UML vers des modèles XML conformes à une certaine DTD. Une première technique consiste à exporter les modèles CPM en modèles XML, puis à transformer un modèle XML vers un autre, en établissant les relations entre les deux DTD correspondantes. Une seconde technique s'appuie sur l'idée qu'un modèle UML outillé peut être interprété comme un système d'information. Ainsi les informations contenues dans les modèles doivent être conservées et synchronisées et un langage permet de faire des requêtes sur ces informations de manière à construire le XML correspondant au modèle UML initial.

Cette approche ne peut pas être utilisée dans notre contexte, car si nous avons bien en sortie des profils dans un format unique (profils au format XML respectant le langage PMDLe), les modèles à convertir sont variés et peuvent être dans différents formats.

Ces différentes approches sont connexes à la nôtre, mais n'y sont pas entièrement adaptées, notamment en raison de la généralité du point de vue que nous adoptons et du fait que nous ne maîtrisons pas les formats des données à traiter.

4. Principe d'intégration de données externes à des profils PMDLe

Notre approche pour l'intégration à des profils PMDLe de données externes issues de profils logiciels consiste à définir puis utiliser des convertisseurs (des parsers) qui permettent de convertir les profils externes selon le formalisme cible.

L'intégration de données externes selon une structure conforme au langage de description de profils se décompose en trois étapes principales (cf. Figure 1) : la *prétraitement* des données permet de spécifier et d'uniformiser l'organisation des données dans les profils (cf. ① sur la Figure 1), la *mise en correspondance des modèles* permet d'associer les éléments des profils externes à la structure de profils cible (cf. ② sur la Figure 1), et l'*exécution du convertisseur* ainsi constitué complète les profils respectant le langage de description de profils (cf. ③ sur la Figure 1).

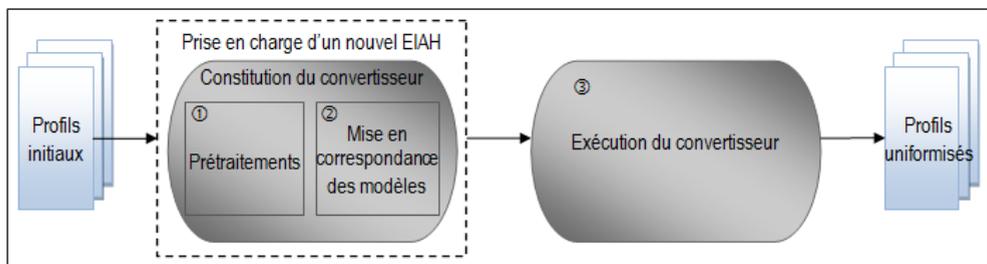


Fig. 1 - Principe du processus d'intégration de données externes à des profils PMDLe.

Le *prétraitement* des profils des EIAH externes concernés (cf. ① de la Figure 1) est un ensemble de transformations portant sur la forme des profils sources visant à représenter ces profils de manière unifiée, ce afin de simplifier la mise en correspondance des modèles.

La *mise en correspondance des modèles* (cf. ② de la Figure 1) a pour but de définir l'alignement de modèles en créant le convertisseur qui sera capable de faire la conversion entre le modèle des profils externes en entrée et celui des structures de profils PMDLe en sortie. Créer un convertisseur nécessite d'associer des éléments du profil source, celui de l'EIAH externe, à des éléments de la structure de profils cible.

L'*exécution du convertisseur* (cf. ③ sur la Figure 1) est le lieu de la conversion de modèles. Elle permet de remplir, grâce au convertisseur créé, la structure de profils avec les données des profils externes pour chaque apprenant, sans que l'utilisateur n'intervienne.

Contrairement à son exécution, la constitution d'un convertisseur, qui correspond à la prise en charge d'un nouvel EIAH, n'est effectuée qu'une fois pour un EIAH donné.

5. L'intégration de données issues de profils externes dans EPROFILEA

Dans cette partie, nous présentons comment nous avons mis en œuvre le principe d'intégration de données externes à des profils PMDLe dans l'environnement EPROFILEA. Nous présentons tout d'abord brièvement l'architecture de l'environnement EPROFILEA, puis nous détaillons le module Tornade qui gère l'intégration de profils logiciels dans cet environnement.

5.1. EPROFILEA - environnement de réutilisation de profils d'apprenants

L'environnement EPROFILEA comporte trois parties : la préparation de profils conformes à l'environnement, c'est-à-dire exprimés selon un même formalisme qui rend possible leur réutilisation (cf. partie gauche de la Figure 2), les éventuelles transformations des profils ainsi unifiés (cf. partie centrale de la Figure 2) et enfin leur exploitation (cf. partie droite de la Figure 2).

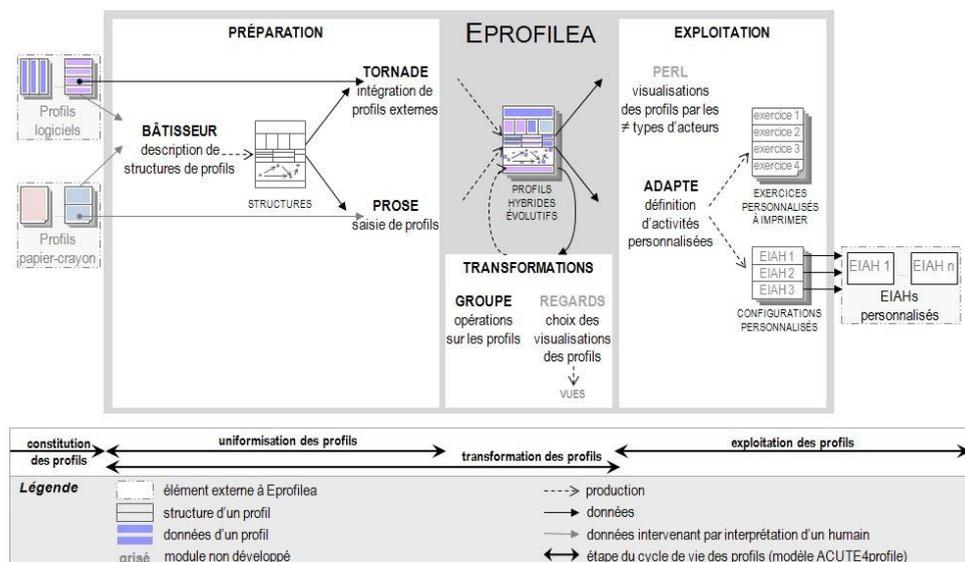


Fig. 2 - Architecture de l'environnement EPROFILEA.

L'uniformisation de la structure de profils hétérogènes est rendue possible par la mise en œuvre du langage de description de profils PMDLe au sein du module Bâtitseur dans EPROFILEA. Cette étape est ensuite complétée par l'intégration des données issues des différents profils dans la structure de profils harmonisée. Dans EPROFILEA, cette intégration se fait avec le module Tornade pour les profils logiciels et avec le module Prose pour les profils papier-crayon. Ainsi, à partir d'une structure de profils créée dans Bâtitseur, Tornade et Prose permettent d'intégrer les données de profils issus respectivement d'EIAH et de profils papier-crayon. À la fin de cette phase de préparation des profils, l'enseignant dispose d'autant de profils que d'apprenants, les profils respectant la structure de profils définie et

donc le formalisme EPROFILEA, ce qui permet de les exploiter dans le reste de l'environnement.

5.2. Tornade - intégration de profils logiciels

Le module Tornade permet à un enseignant, à l'aide de convertisseurs (des parsers que nous appelons *tourbillons*), d'intégrer aux structures de profils qu'il a définies dans Bâtitteur, les données de chacun de ses élèves contenues dans des profils logiciels, complétant ainsi le cas échéant les profils déjà en partie renseignés dans Prose. Un tourbillon, noté $T(EIAH_\alpha, StrPro_\beta)$, propre au couple formé de l' $EIAH_\alpha$ et de la structure de profils EPROFILEA $StrPro_\beta$, est un adaptateur qui permet de transformer tout ou partie du modèle de profils de l' $EIAH_\alpha$ en un ensemble de briques contenues dans la structure de profils $StrPro_\beta$. Un tourbillon est défini pour un $EIAH_\alpha$ une fois pour toute par un expert à l'aide du module Tornade, puis utilisé par un enseignant aussi souvent que nécessaire pour remplir automatiquement les profils EPROFILEA des apprenants avec les données contenues dans leur profil de l' $EIAH_\alpha$ selon la structure de profils $StrPro_\beta$.

5.2.1. Principe de fonctionnement du module Tornade

Dans le module Tornade de l'environnement EPROFILEA (cf. Figure 1), l'intégration de données externes selon une structure conforme au formalisme EPROFILEA se décompose en deux étapes principales (cf. Figure 3) : la constitution d'un convertisseur (le tourbillon) et son exécution. La constitution du tourbillon comporte des *prétraitements* sur les données pour spécifier et uniformiser l'organisation des données dans les profils et la *production d'un tourbillon* elle-même qui permet d'associer les éléments des profils externes aux éléments d'une structure de profils donnée. L'*exécution d'un tourbillon* complète les profils EPROFILEA associés à la structure de profils donnée.

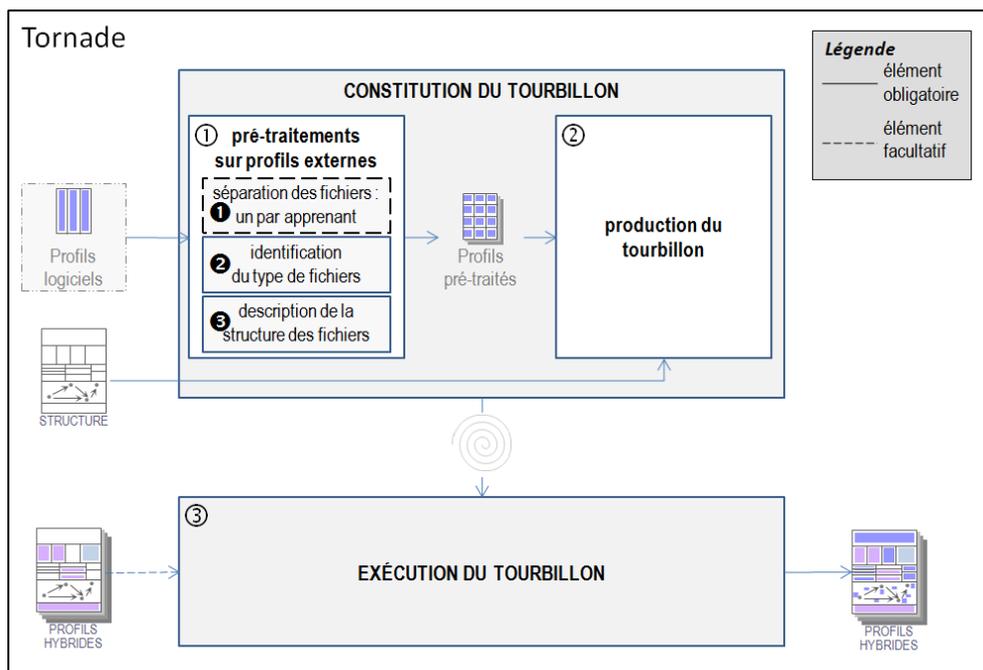


Fig. 3 - Architecture du module Tornade.

Le *prétraitement* des profils des EIAH externes concernés (cf. ① de la Figure 3) consiste tout d'abord à séparer les fichiers externes de telle manière qu'il y ait autant de fichiers que d'apprenants (① de la Figure 3 et Figure 4), puis à identifier d'une part le type de fichiers (② de la Figure 3 et Figure 5) et d'autre part l'organisation des données de l'apprenant dans un fichier (③ de la Figure 3 et Figure 6).

La séparation des fichiers externes (① de la Figure 3) n'est nécessaire que si l'EIAH externe fournit un ou plusieurs fichiers contenant chacun les données de plusieurs apprenants. Lors de l'identification de l'organisation des données de l'apprenant (③ de la Figure 3), l'expert décrit les données contenues dans les profils, ce qui permet à Tornade de fractionner chaque profil de manière à organiser les données qu'il contient dans un tableau. Cette description de l'organisation des profils du logiciel externe est établie une seule fois par EIAH.

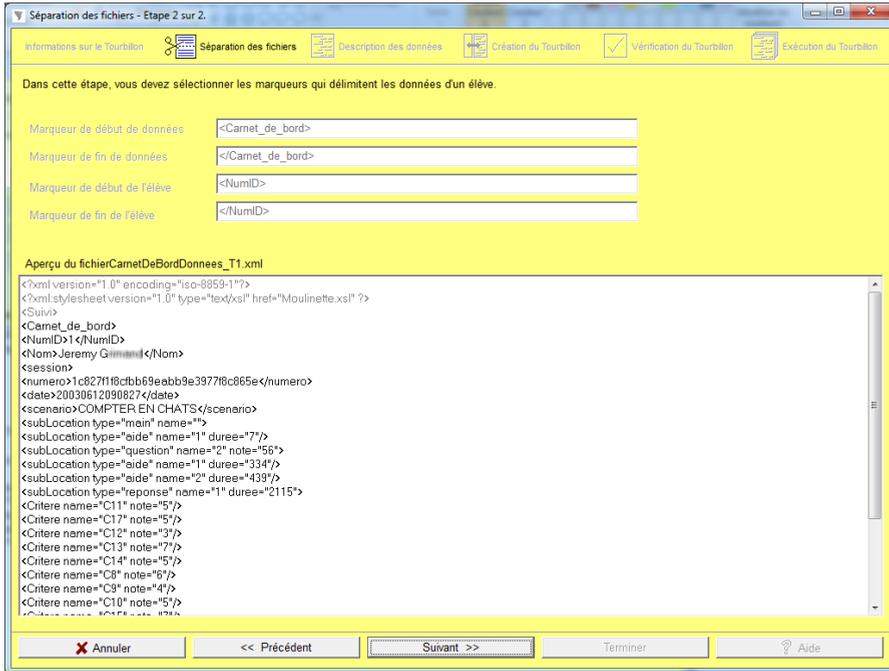


Fig. 4 - Prétraitement des profils avec Tornade – séparation d'un fichier unique.

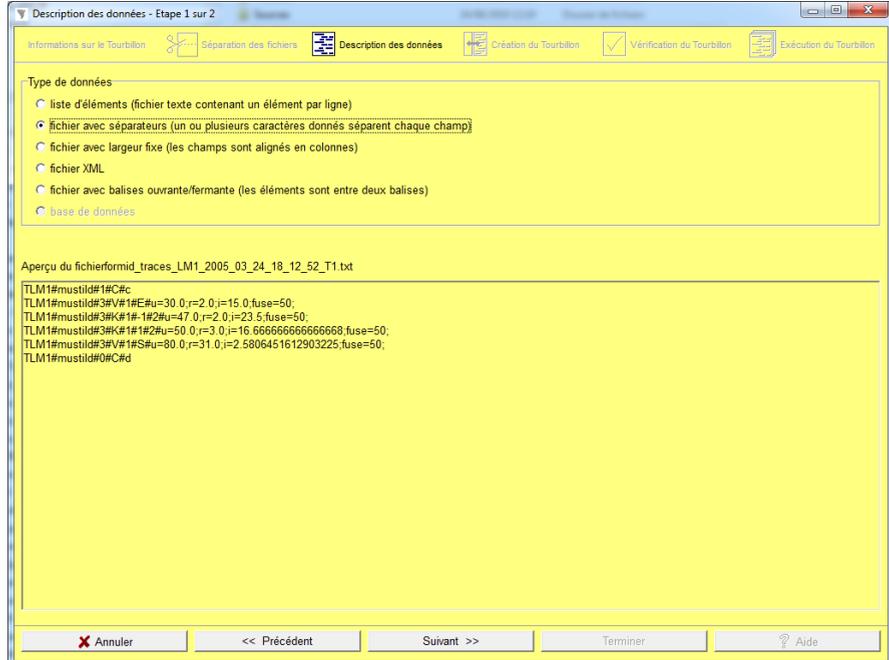


Fig. 5 - Prétraitement des profils avec Tornade – identification du type de fichiers.

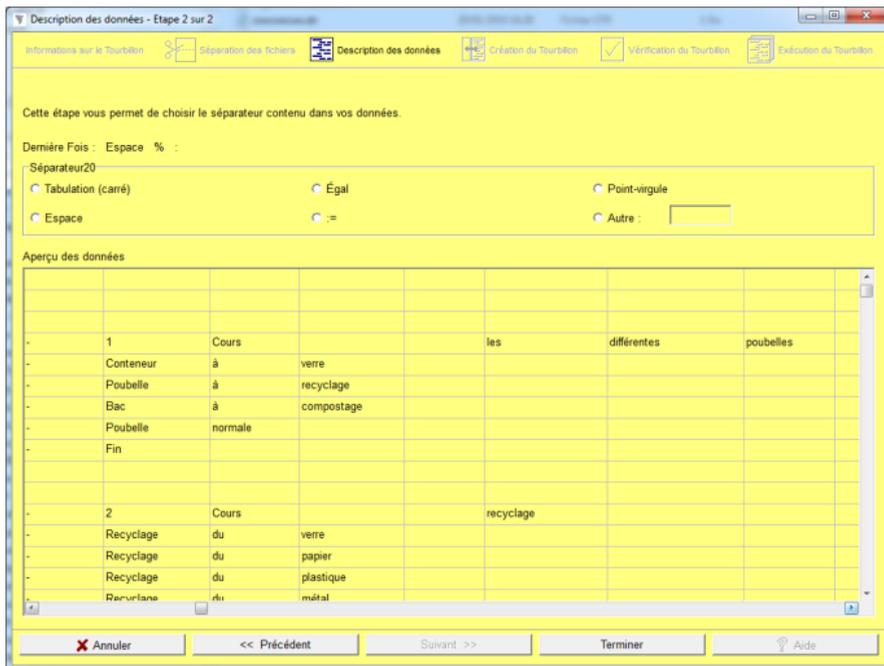


Fig.- 6 - Prétraitement des profils avec Tornado – description des données.

La *production du tourbillon* (cf. ② de la Figure 3 et Figure 7) a pour but de définir l'alignement de modèles en créant le convertisseur (le tourbillon) qui sera capable de faire la conversion entre le modèle des profils externes en entrée et celui des structures de profils EPROFILEA en sortie. Plusieurs briques de cette structure de profils EPROFILEA peuvent concerner l'EIAH externe dont les données doivent être intégrées. Ce sont ces différentes briques que le tourbillon à créer doit savoir remplir dynamiquement. Le traitement d'une brique est générique et s'adapte automatiquement à tous les types de briques conformes au formalisme EPROFILEA opérationnalisant le langage PMDLe. Si la structure de profils est une modification d'une structure déjà traitée, Tornado s'appuie sur le premier tourbillon $T(EIAH_{\alpha}, StrPro_{1a})$ pour créer le second $T(EIAH_{\alpha}, StrPro_{1b})$. Pour créer un tourbillon, Tornado associe des éléments du profil source, celui de l'EIAH externe, à des éléments de la structure de profils cible (cf. Figure 7), en appliquant d'une part des conversions d'échelles dans le cas où l'échelle des données du profil source ne correspond pas à l'échelle de la structure de profils cible, et d'autre part des opérations, pour permettre d'associer plusieurs éléments du profil source à un unique élément de la structure de profils cible.

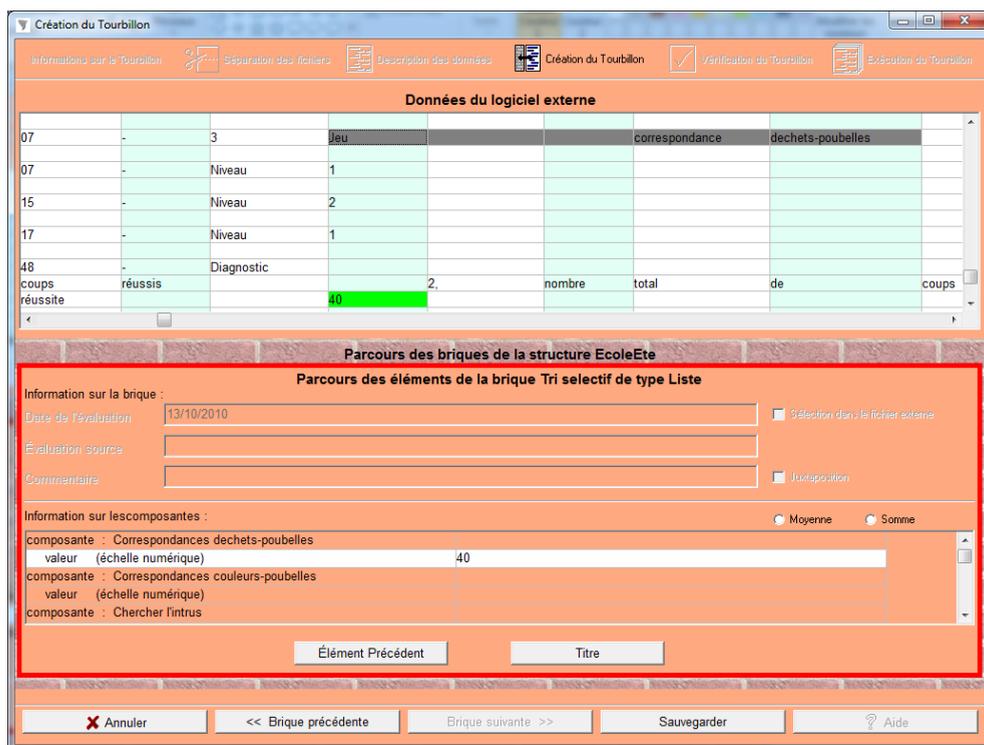


Fig. 7 - Production d'un tourbillon avec Tornade.

L'exécution du tourbillon (cf. ③ de la Figure 3 et Figure 9) est le lieu de la conversion de modèles. Elle permet de remplir, grâce au convertisseur créé (le tourbillon), la structure de profils avec les données des profils externes sans que l'utilisateur n'intervienne. Tornade remplit la structure de profils dynamiquement pour chaque apprenant, vérifiant qu'il ne manque aucun apprenant (cf. Figure 9) et que la création des profils EPROFILEA, ou la complétion dans le cas des profils hybrides, se fait sans erreur (cf. Figure 8).

Pour résumer, lors de l'utilisation de Tornade :

- ① la phase de prétraitement des données n'est faite qu'une seule fois par EIAH par un expert ;
- ② la phase de constitution du tourbillon n'est faite qu'une seule fois pour un couple EIAH / structure de profils par un expert ;
- ③ la phase d'exécution du tourbillon sera effectuée par un enseignant chaque fois qu'il voudra créer ou mettre à jour ses profils.

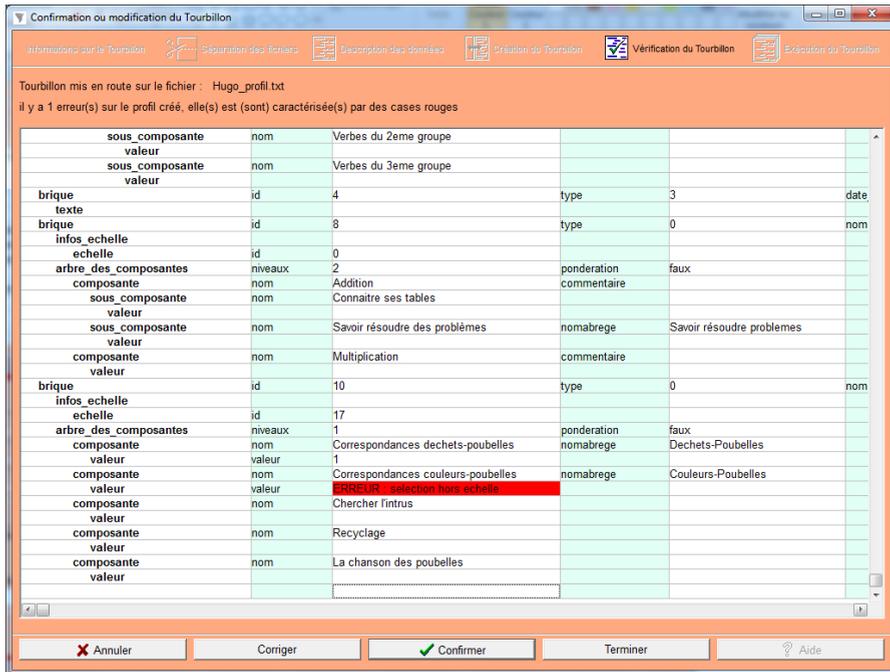


Fig. 8 - Vérification d'un tourbillon avec Tornado.

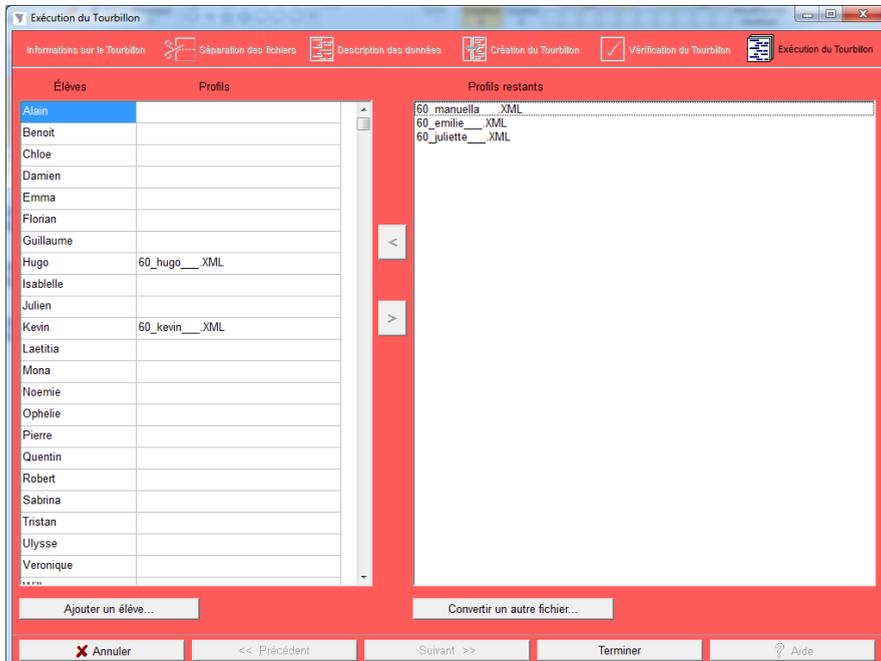


Fig. 9 - Exécution d'un tourbillon avec Tornado.

5.2.2. Exemple d'utilisation du module Tornade

Pour illustrer le fonctionnement du module Tornade et de l'étape de conversion des modèles de profils, prenons l'exemple de deux profils issus du logiciel TRI (Jean-Daubias 2007) : celui d'Alain et celui de Benoît (cf. Figure 10). Lors de l'utilisation du logiciel, chaque élève a fait une série d'exercices différents, mais tous deux ont joué au jeu « Chercher l'intrus ».

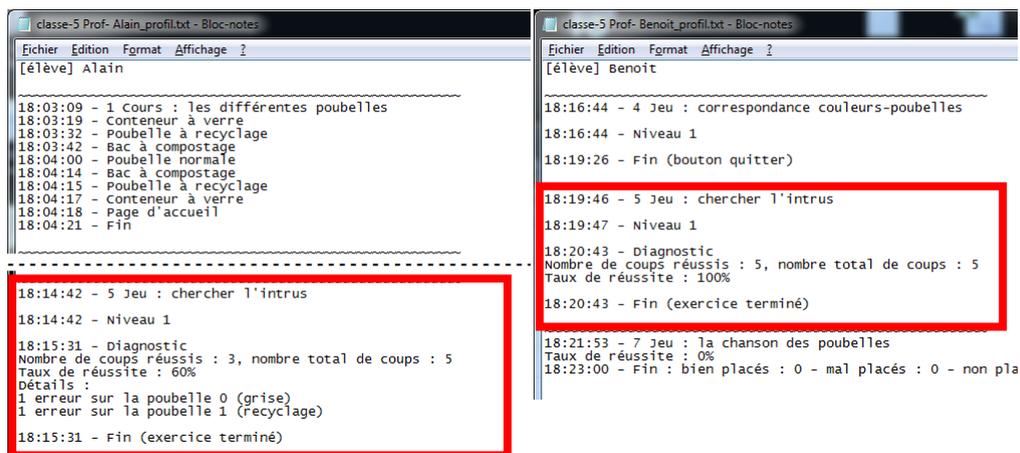


Fig. 10 – Profils issus du logiciel TRI.

```

- <structure id="60" nom="EcoleEle" createur="Marie" date_creation="28/05/2007"
date_derniere_modif="11/01/2010" nom_eleve="" prenom_eleve=""
+ <brique id="16" type="0" nom="Francais" indice="0"></brique>
+ <brique id="17" type="3" indice="1"></brique>
+ <brique id="15" type="0" nom="Mathématiques" indice="2"></brique>
- <brique id="18" type="0" nom="Tri selectif" logiciel_externe="Logiciel de tn selectif" indice="3">
+ <infos_echelle></infos_echelle>
- <arbre_des_composantes niveaux="1" ponderation="faux">
- <composante nom="Correspondances dechets-poubelles" nomabrege="Dechets-Poubelles">
<valeur_id_echelle="17" num="0"/>
</composante>
- <composante nom="Correspondances couleurs-poubelles" nomabrege="Couleurs-Poubelles">
<valeur_id_echelle="17" num="0"/>
</composante>
- <composante nom="Chercher l'intrus">
<valeur_id_echelle="17" num="0"/>
</composante>
- <composante nom="Recyclage">
<valeur_id_echelle="17" num="0"/>
</composante>
- <composante nom="La chanson des poubelles">
<valeur_id_echelle="17" num="0"/>
</composante>
</arbre_des_composantes>
</brique>
</structure>

```

Fig. 11 – Structure de profil EPROFILEA.

Afin d'intégrer leurs résultats dans un profil dont la structure respecte le formalisme PMDL (cf. Figure 11), Tornade va dans un premier temps demander à l'expert d'identifier le type de fichier (cf. Figures 5 et 6) : dans le cas du logiciel

TRI, il y a un fichier par élève et chaque fichier les valeurs sont associées à des champs séparés par les caractères espace (#32) et « % » (cf. partie gauche de la Figure 12).

```

<ListeLogicielsExternes>
+ <Logiciel nom="Carnet de Bord"></Logiciel>
+ <Logiciel nom="Formid"></Logiciel>
<Logiciel nom="Pépète"></Logiciel>
- <Logiciel nom="Logiciel de tri sélectif"
  <ImportationFichier>
  <TypeFichier>1</TypeFichier>
  <Construction>
  <Separateur>#32</Separateur>
  <Separateur>%</Separateur>
  </Construction>
  </ImportationFichier>
</Logiciel>
+ <Logiciel nom="Ambre"></Logiciel>
<Logiciel nom="Géonote"></Logiciel>
+ <Logiciel nom="Metadyne"></Logiciel>
<Logiciel nom="MoreMaths"></Logiciel>
</ListeLogicielsExternes>
- <structure id="60" nom="EcoleEte" createur="Marie" date_creation="28/05/2007"
  date_derniere_modif="01/02/2011" nom_eleve="" prenom_eleve="Alain">
  prenom_eleve="[élève][04][014]0[024][003]0[013][004]1[014]0[024]"
  + <brique id="16" type="0" nom="Français" indice="0"></brique>
  + <brique id="17" type="3" indice="1"></brique>
  + <brique id="15" type="0" nom="Mathématiques" indice="2"></brique>
  - <brique id="18" type="0" nom="Tri sélectif" logiciel_externe="Logiciel de tri sélectif" indice="3">
  + <infos_echelle></infos_echelle>
  - <arbre_des_composantes niveaux="1" ponderation="faux">
  + <composante nom="Correspondances dechets-poubelles" nomabrege="Dechets-Poubelles">
  </composante>
  + <composante nom="Correspondances couleurs-poubelles" nomabrege="Couleurs-Poubelles">
  </composante>
  - <composante nom="Chercher l'intrus">
  - <valeur id_echelle="17" num="0">
  - <evaluation date="2011/02/01" source="">
  - <unique>
  5Jeu.chercherl'intrus[04]2[014]86[024][003]4[013][004]2[014]6[024]
  - <conversion>
  <num_num BornelInf="0" BornelInitiale="0" BorneSup="100" pas="0,01"
  PasInitial="0,01"/>
  </conversion>
  </unique>
  </evaluation>
  </valeur>
  </composante>
  + <composante nom="Recyclage"></composante>
  + <composante nom="La chanson des poubelles"></composante>
  </arbre_des_composantes>
  </brique>
</structure>

```

Fig. 12 – Fichiers « tourbillon » créés par le module Tornade.

```

- <structure id="60" nom="EcoleEte" createur="Marie" date_creation="01/02/2011"
  date_derniere_modif="01/02/2011" nom_eleve="" prenom_eleve="Alain">
  + <brique id="16" type="0" nom="Français" indice="0"></brique>
  + <brique id="17" type="3" indice="1"></brique>
  + <brique id="15" type="0" nom="Mathématiques" indice="2"></brique>
  - <brique id="18" type="0" nom="Tri sélectif" logiciel_externe="Logiciel de tri sé
  + <infos_echelle></infos_echelle>
  - <arbre_des_composantes niveaux="1" ponderation="faux">
  + <composante nom="Correspondances dechets-poubelles" nomabrege="
  </composante>
  + <composante nom="Correspondances couleurs-poubelles" nomabrege="
  </composante>
  - <composante nom="Chercher l'intrus">
  - <valeur id_echelle="17" num="0">
  <evaluation date="2011/02/01" source="">60 </evaluation>
  </valeur>
  </composante>
  + <composante nom="Recyclage"></composante>
  + <composante nom="La chanson des poubelles"></composante>
  </arbre_des_composantes>
  </brique>
</structure>
- <structure id="60" nom="EcoleEte" createur="Marie" date_creation="01/02/2011"
  date_derniere_modif="01/02/2011" nom_eleve="" prenom_eleve="Benoit">
  + <brique id="16" type="0" nom="Français" indice="0"></brique>
  + <brique id="17" type="3" indice="1"></brique>
  + <brique id="15" type="0" nom="Mathématiques" indice="2"></brique>
  - <brique id="18" type="0" nom="Tri sélectif" logiciel_externe="Logiciel de tri sé
  + <infos_echelle></infos_echelle>
  - <arbre_des_composantes niveaux="1" ponderation="faux">
  + <composante nom="Correspondances dechets-poubelles" nomabrege="
  </composante>
  + <composante nom="Correspondances couleurs-poubelles" nomabrege="
  </composante>
  - <composante nom="Chercher l'intrus">
  - <valeur id_echelle="17" num="0">
  <evaluation date="2011/02/01" source="">100 </evaluation>
  </valeur>
  </composante>
  + <composante nom="Recyclage"></composante>
  + <composante nom="La chanson des poubelles"></composante>
  </arbre_des_composantes>
  </brique>
</structure>

```

Fig. 13 – Profils EPROFILEA complétés par le module Tornade.

L'expert va ensuite spécifier dans Tornade les liens existants entre les champs des fichiers issus du logiciel TRI et ceux du profil EPROFILEA. Ces liens entre les modèles de profils sont mémorisés dans le tourbillon sous forme d'un ensemble de codes (cf. partie droite de la Figure 12).

Pour compléter les profils EPROFILEA, l'enseignant n'aura plus qu'à lancer le tourbillon afin que Tornade convertisse automatiquement les informations contenues dans les profils externes (cf. Figure 13).

5.2.3. Évaluation du module Tornade

Nous avons testé le module Tornade avec les fichiers de sortie variés (des profils d'apprenants) de 14 logiciels pédagogiques différents non conçus pour fonctionner avec l'environnement EPROFILEA. Les tests ont montré que Tornade permet d'intégrer les données contenues dans les profils correspondants.

De plus, Tornade permet de faire des opérations (sommations ou moyennes) sur les données importées des profils externes au moment de l'intégration, ce qui permet par exemple de placer dans un élément du profil la somme de différentes valeurs présentes à des endroits différents dans le profil initial. Nous devons à présent enrichir le système de nouvelles fonctionnalités pour permettre de traiter plus facilement certains profils, comme la recherche dans un profil de toutes les occurrences d'une compétence donnée ou des opérations plus complexes sur les données importées.

Par ailleurs, une enseignante de collège a utilisé avec succès le logiciel pour convertir 12 profils issus d'un EIAH externe dans une structure de profils préalablement créée, ce en 10 minutes. Nous avons également présenté le module Tornade à deux enseignants de primaire. Ils en ont compris le principe, mais ont insisté sur le fait que la majorité des enseignants ne sont pas prêts à utiliser de tels outils. Nous sommes donc conscientes que ce module, pour sa partie constitution de tourbillons, contrairement au reste de l'environnement EPROFILEA, doit être utilisé par des experts, qui peuvent être les conseillers informatique de l'établissement scolaire ou des enseignants ayant suffisamment de recul sur les logiciels utilisés et un minimum de compétences en informatique. Ces experts fabriqueront les tourbillons correspondant aux besoins et aux habitudes de travail des enseignants. Les enseignants pourront ensuite simplement exécuter les tourbillons dans Tornade, sans avoir à les modifier, pour créer ou mettre à jour les profils de leurs élèves.

En complément, nous avons effectué un test de la partie constitution des convertisseurs de Tornade avec une enseignante de Français de Collège « non-experte ». Notre objectif était d'évaluer les difficultés rencontrées par un enseignant non-expert pour créer un tourbillon. Nous avons pu constater qu'avec un guidage modéré de la part d'une des expérimentatrices, l'enseignante a pu créer un tourbillon adapté à la structure de profils qu'elle avait préalablement créée et à l'EIAH concerné, ce en 50 minutes. Ces observations sont encourageantes pour l'utilisation de Tornade par des enseignants-experts.

6. Conclusion

Dans cet article, nous avons présenté le processus d'intégration de données hétérogènes que nous proposons pour traiter notre problématique de réutilisation de profils d'apprenants issus de sources hétérogènes dont le formalisme de représentation n'est pas connu. Ce processus comporte deux phases distinctes : l'initialisation qui permet la prise en charge de profils issus d'EIAH jusqu'alors non gérés, et l'intégration des données à proprement parler qui consiste à convertir les profils externes pour les intégrer dans un profil respectant le formalisme cible.

Nous avons mis en œuvre ce processus au sein de l'environnement EPROFILEA pour l'intégration des données de profils externes aux structures de profils respectant le formalisme basé sur le langage de modélisation de profils PMDL. La démarche adoptée est générique et le processus est adapté à tous les éléments respectant PMDL. Combiné à Prose, le module Tornade permet de créer dans l'environnement des profils hybrides, combinant dans un même profil des informations issus de sources variées, tant logicielles que papier-crayon. Les profils ainsi uniformisés peuvent ensuite être exploités par les autres modules d'EPROFILEA : visualisations de profils et propositions d'activités pédagogiques personnalisées selon le contenu des profils (Lefevre *et al.* 2009). L'intégration de données hétérogènes est en effet l'une des étapes préparatoires aux exploitations, qui sont la finalité de l'environnement EPROFILEA.

Références

HACID M.-S. & REYNAUD C. (2004). L'intégration de sources de données. *Revue I3 (Information Interaction Intelligence)*, vol. 4, n°2.

JEAN-DAUBIAS S. (2007) Logiciel Tri, <http://liris.cnrs.fr/stephanie.jean-daubias/projets/logiciel-TriSelectif.html>.

JEAN-DAUBIAS S. & GINON B. (2010). Des profils d'apprenant évolutifs. *TICE 2010*, Nancy, France, ISBN 2-9516740-5-8.

JEAN-DAUBIAS S., EYSSAUTIER-BAVAY C., LEFEVRE M. (2009). Modèles et outils pour rendre possible la réutilisation informatique de profils d'apprenants hétérogènes. *STICEF*, vol. 16, ISSN 1764-7223.

LAFORCADE P. (2005). Approche par transformation de modèles pour la conception d'EIAH. *EAIH'2005*. Montpellier, France, p. 213-224.

LEFEVRE M., CORDIER A., JEAN-DAUBIAS S., GUIN N. (2009). A Teacher-dedicated Tool Supporting Personalization of Activities. *ED-MEDIA 2009*, Honolulu, 22-26 juin 2009, p.1136-1141.

WIEDERHOLD G. (1992). Mediators in the architecture of future information systems. *Computer*, vol. 25, n°3, p.38-49.