

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

Marie Lefevre et  
**Stéphanie Jean-Daubias**  
LIRIS-Silex

# Préalable

- Profil d'apprenant
  - ensemble d'informations
  - concernant l'apprentissage d'un apprenant
    - connaissances (ex.  $a+a=2a$ ), connaissances erronées (ex.  $a+a=a^2$ )
    - compétences (ex. développer une expression)
    - comportement (ex. travail par essais-erreurs)
  - valuées selon une échelle
    - numériques (ex. 12/20, 7/8, -2)
    - textuelles graduées (ex : B, partiellement maîtrisé, (2) Expliciter)
    - textuelles non graduées : (ex. autonome, attentif)
  - collectées ou déduites à l'issue d'activités pédagogiques
    - informatisées
    - papier-crayon
- Attention
  - profils : constitués en amont dans une phase de diagnostic
  - profils  $\neq$  profils-types ou stéréotypes

Juin 2000

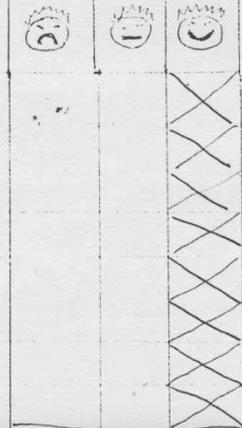
Atteint Vitesse suff	Maîtrise règles	Utilise outils	Identifie Nombres	Identifie fonctions	Utilise verbes	Utilise diction.	Compte opération	Problème
4	3	3	2	3	3	1	3	4
2	3	3	2	3	3	2	2	4

Martin

Emilie V

Bilan Mathématiques

- lire les nombres naturels jusqu'à 9999
- écrire les nombres naturels jusqu'à 9999
- produire différentes écritures d'un même nombre naturel
- ranger des nombres naturels (ordre croissant ou décroissant)
- additionner entiers
- trouver la différence de 2 entiers
- multiplier 2 entiers avec tous les chiffres différents de zéro



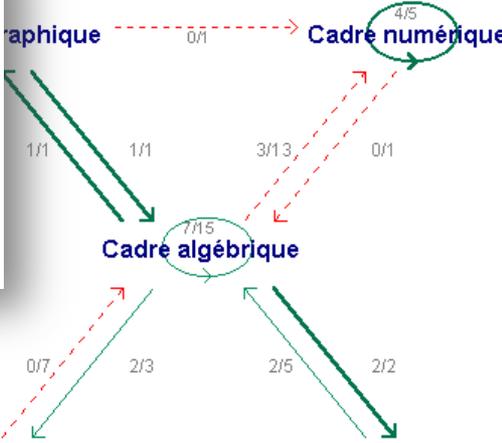
You Can Do.....

subtractions like these.

KEY



<p>23-12</p> <p>Two digit subtract two digit (no adjustment)</p>	<p>76-28</p> <p>Two digit subtract two digit (adjustment from tens to units)</p>	<p>459-234</p> <p>Three digit subtract three digit (no adjustment)</p>	<p>574-359</p> <p>Three digit subtract three digit (adjustment tens to units)</p>	<p>364-17</p> <p>Three digit subtract three digit (adjustment hundreds to tens to units)</p>
--	--	--	---	--



- Je pense que je sais la technique de la multiplication.
- Je pense que je sais un peu la technique de la multiplication pour les nombres décimaux.
- Je pense que je sais à quel moment utiliser la multiplication.

Conversion

- Je pense que je ne sais pas utiliser les unités de masse.

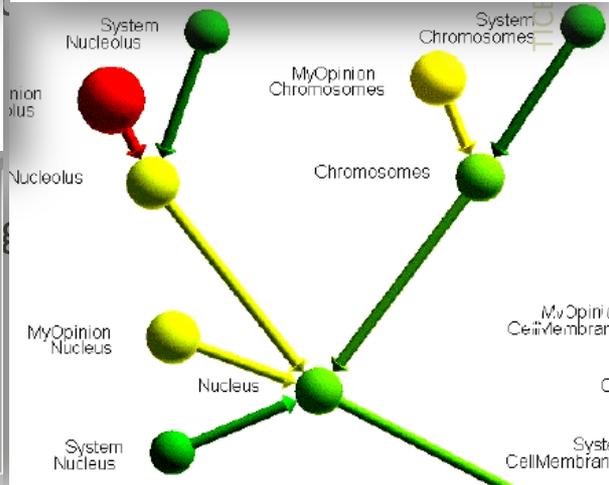
TOPIC	% COVERED	% KNOWN	known	problematic	mis conceptions	topic size
Computer Basics	50	25				
Programming Basics	50	25				

Calcul algébrique

Bonne maîtrise des règles de transformation	13	
Maîtrise technique fragile		
- Règles de transformation non maîtrisées, mais identification correcte du rôle des opérateurs + et ×		
Utilisation inadaptée des parenthèses qui conduit à un résultat correct	1	
Utilisation inadaptée des parenthèses qui conduit à un résultat incorrect		
Utilisation de règles de transformation fausses identifiées	3	
Erreurs de signes au cours d'un calcul		
+ Identification incorrecte du rôle des opérateurs + et ×	2	
Calcul sans repérage des blocs		
Calcul non identifié	1	

Langage naturel

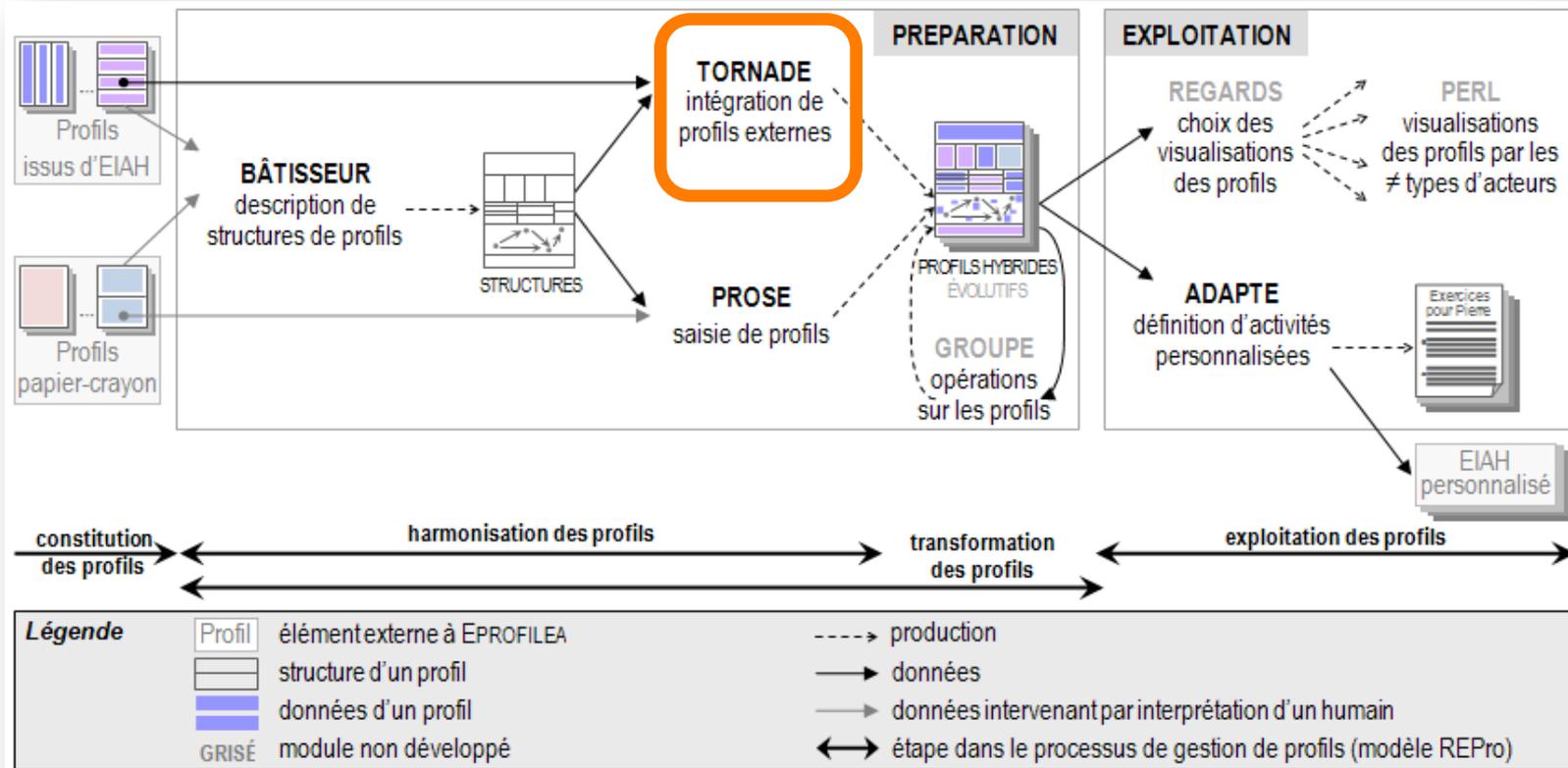
Cadre géométrique



# Contexte : le projet PERLEA

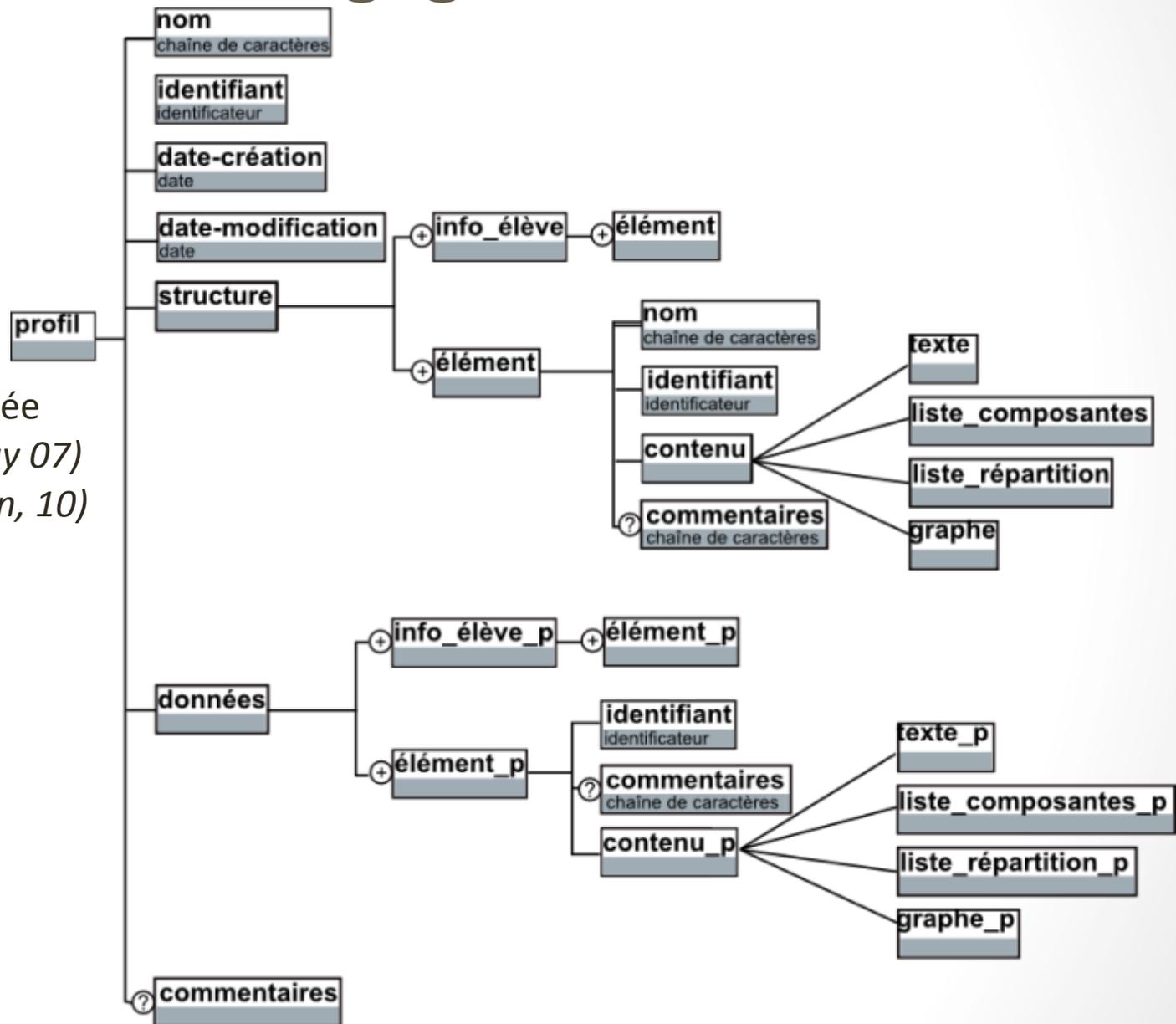
- Proposer des outils d'aide à l'enseignant
  - réutilisation de profils d'apprenants hétérogènes existants
  - exploitations mutualisées personnalisées en classe
- Démarche générique / environnement unifié
  - éléments externes variés
    - profils d'apprenants
    - activités pédagogiques
  - existants ou à venir
  - papier-crayon ou logiciels
  - différents niveaux
  - différentes disciplines

# Contexte : l'environnement EPROFILEA



# Contexte : le langage PMDLe

Profiles  
MoDeling  
Language –  
evolution :  
structuration  
générale simplifiée  
(Eyssautier-Bavay 07)  
(J-Daubias, Ginon, 10)



# Problématique

- Intégration de données issues de sources hétérogènes

- sources

- profils d'apprenant informatisés
- différents formats et structures de stockage
- modèle de structure pas forcément connu

- résultat

- combinaison de plusieurs sources
- exploitation que d'une partie des données
- structure respectant un formalisme donné

⇒ Approche générique

- Problème informatique classique

- recherche d'informations, aide à la décision, web sémantique...

**Bilan Mathématiques**

1 - lire les nombres naturels jusqu'à 9999  
3 - écrire les nombres naturels jusqu'à 9999  
5 - produire différentes écritures d'un même nombre naturel  
6 - ranger des nombres naturels (ordre croissant ou décroissant)  
20 - additionner entiers  
21 - trouver la différence de 2 entiers  
23 - multiplier 2 entiers avec tous les chiffres différents de zéro

**You Can Do...**

subtractions like these. **KEY**

29-12 Two digit subtract two digit (adjustment)  
76-28 Two digit subtract two digit (adjustment and tens to units)  
459-234 Three digit subtract three digit (adjustment)  
674-350 Three digit subtract three digit (adjustment tens to units)  
884-175 Three digit subtract three digit (adjustment hundreds to tens and tens to units)

TOPIC	% COVERED	% KNOWN	known	problematic	mis conceptions
Computer Basics	50	25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Programming Basics	50	25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

System  
Chromosomes  
Nucleus  
Mitochondria  
Cytoplasm

# État de l'art

- Systèmes médiateurs

- exécution d'une requête sur l'ensemble des sources distribuées

- EPROFILEA [
- les modèles sources ne sont pas forcément connus à l'avance
  - les modèles sources existent déjà

- Entrepôts de données

- intégrer et stocker toutes les données dans un environnement

- EPROFILEA [
- interrogation ponctuelle des données (en fonction des besoins de l'enseignant, au fur et à mesure de la création ou mäj des profils)

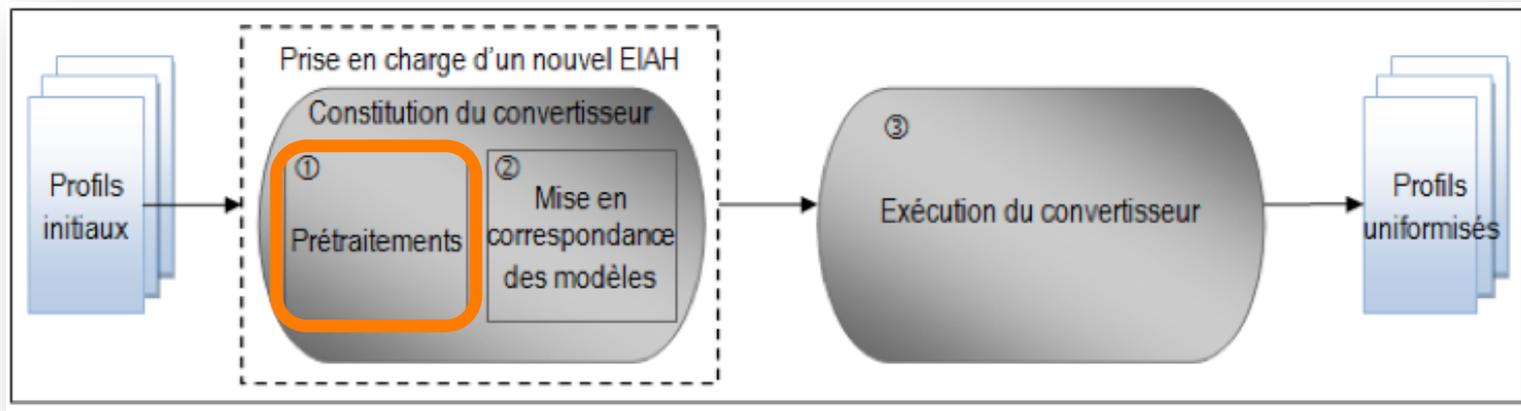
- En EIAH

- transformation de modèles *ex UML -> XML (Laforcade 05)*

- profils d'utilisateur ubiquitaire *ex (Berkovsky et al. 05)*

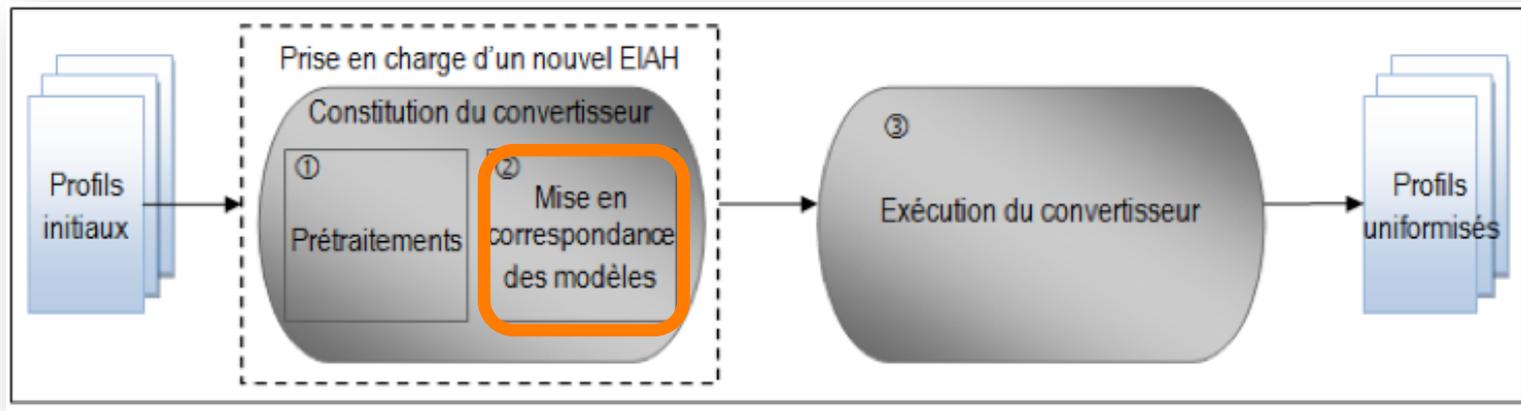
- EPROFILEA [
- les modèles des profils sources ne sont pas connus à l'avance

# Principe d'intégration de données externes à des profils PMDLe



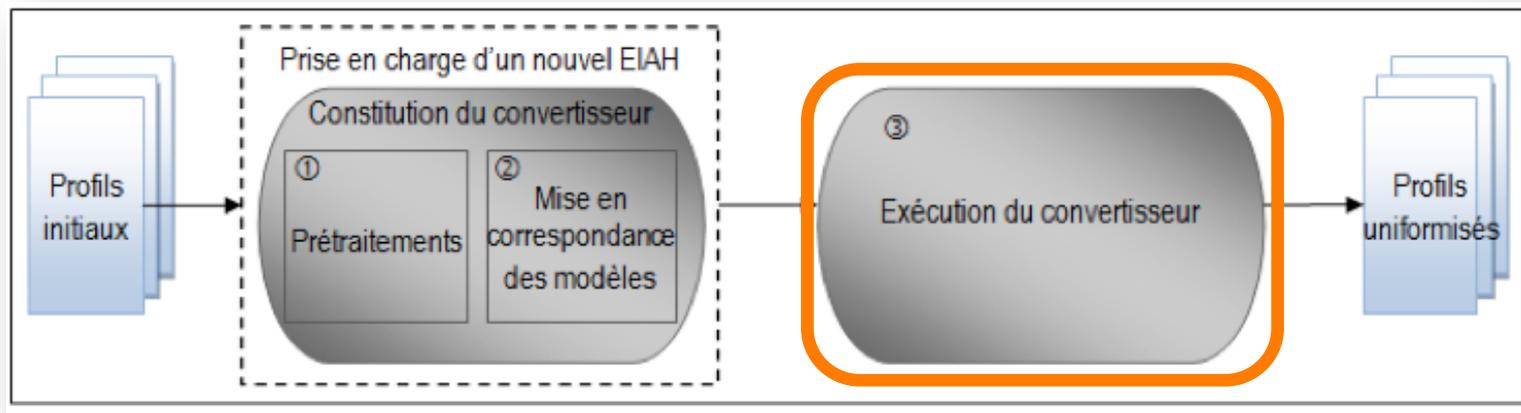
- Prétraitement des profils des EIAH externes
  - but : simplifier la mise en correspondance des modèles
  - ensemble de transformations
  - portant sur la forme des profils sources
  - visant à représenter ces profils de manière unifiée

# Principe d'intégration de données externes à des profils PMDLe



- Mise en correspondance des modèles
  - but : alignement de modèles
  - pour chaque profil en entrée, création d'un convertisseur
    - capable de faire la conversion entre le modèle des profils externes et celui des structures de profils PMDLe
    - association des éléments du profil source à des éléments de la structure de profils cible

# Principe d'intégration de données externes à des profils PMDLe

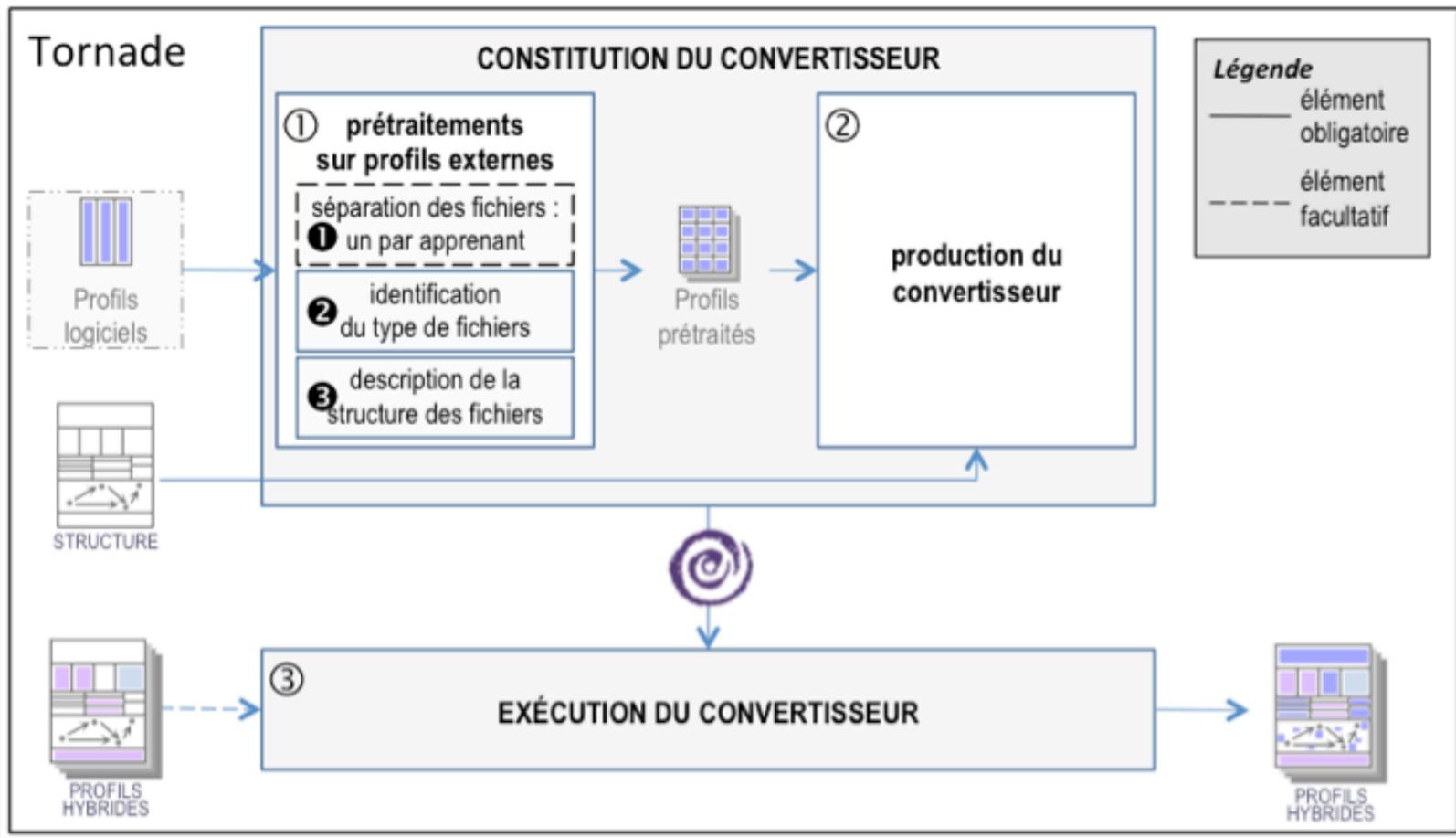


- Exécution du convertisseur
  - but : remplir les profils
  - lieu de la conversion de modèles
  - permet de remplir la structure de profils
    - avec les données des profils externes
    - pour chaque apprenant
    - sans que l'utilisateur n'intervienne

# Intégration de données issues de profils externes dans Eprofilea

- Le module Tornade
  - permet à un enseignant non informaticien
  - d'intégrer les données issues de sources diverses
  - dans une structure de profils précédemment créée
  - pour obtenir un profil par apprenant
- Initialisation par un expert
  - pour une structure de profils pmdle
  - définition de convertisseurs
    - un par source de données

# Le module Tornade



# Prétraitement des profils : Séparation d'un fichier unique

The image shows the 'Tornado' software interface. On the left, a flowchart titled 'CONSTITUTION DU CONVERTISSEUR' outlines three steps: 1. prétraitements sur profils externes, 2. identification du type de fichiers, and 3. description de la structure des fichiers. Step 1 is highlighted with an orange box and contains the text 'séparation des fichiers : 1 un par apprenant'. Below this, 'PROFILS LOGICIELS' and 'PROFILS HYBRIDES' are shown with arrows pointing to the process. On the right, a dialog box titled 'Séparation des fichiers - Etape 2 sur 2' is open. It contains a toolbar with icons for 'Séparation des fichiers', 'Description des données', 'Création du Tourbillon', 'Vérification du Tourbillon', and 'Exécution du Tourbillon'. The main area of the dialog is yellow and contains the instruction: 'Dans cette étape, vous devez sélectionner les marqueurs qui délimitent les données d'un élève.' Below this, there are four input fields for delimiters: 'Marqueur de début de données' (containing '<Carnet\_de\_bord>'), 'Marqueur de fin de données' (containing '</Carnet\_de\_bord>'), 'Marqueur de début de l'élève' (containing '<NumID>'), and 'Marqueur de fin de l'élève' (containing '</NumID>'). At the bottom, there is a preview of an XML file named 'CarnetDeBordDonnees\_T1.xml' with the following content:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml:stylesheet version="1.0" type="text/xsl" href="Moulinette.xsl" ?>
<Suivi>
<Carnet_de_bord>
<NumID>1</NumID>
<Nom>Jeremy Gimard</Nom>
<session>
<numero>1c827f1f8cfbb69eabb9e3977f8c865e</numero>
<date>20030612090827</date>
<scenario>COMPTER EN CHATS</scenario>
<subLocation type="main" name="">
<subLocation type="aide" name="1" duree="7"/>
<subLocation type="question" name="2" note="56">
<subLocation type="aide" name="1" duree="334"/>
<subLocation type="aide" name="2" duree="439"/>
<subLocation type="reponse" name="1" duree="2115">
<Critere name="C11" note="5"/>
<Critere name="C17" note="5"/>
<Critere name="C12" note="3"/>
<Critere name="C13" note="7"/>
<Critere name="C14" note="5"/>
<Critere name="C8" note="6"/>
<Critere name="C9" note="4"/>
<Critere name="C10" note="5"/>
<Critere name="C11" note="5"/>
```

At the bottom of the dialog, there are buttons for 'Annuler', '<< Précédent', 'Suivant >>', 'Terminer', and '? Aide'.

# Prétraitement des profils : Identification du type de fichier

The image shows the 'Tornado' software interface. On the left, a workflow diagram titled 'CONSTITUTION DU CONVERTISSEUR' is visible. It includes three main steps: 1. 'prétraitements sur profils externes' (subdivided into 'séparation des fichiers : 1 un par apprenant', '2 identification du type de fichiers', and '3 description de la structure des fichiers'), and 3. 'EXÉCUTION'. On the right, a dialog box titled 'Description des données - Etape 1 sur 2' is open. It features a toolbar with icons for 'Séparation des fichiers', 'Description des données', 'Création du Tourbillon', 'Vérification du Tourbillon', and 'Exécution du Tourbillon'. The 'Description des données' step is active. Below the toolbar, the 'Type de données' section contains several radio button options: 'liste d'éléments (fichier texte contenant un élément par ligne)', 'fichier avec séparateurs (un ou plusieurs caractères donnés séparent chaque champ)' (which is selected), 'fichier avec largeur fixe (les champs sont alignés en colonnes)', 'fichier XML', 'fichier avec balises ouvrante/fermante (les éléments sont entre deux balises)', and 'base de données'. At the bottom of the dialog, an 'Aperçu du fichier' section shows a preview of the data from 'fichierformid\_traces\_LM1\_2005\_03\_24\_18\_12\_52\_T1.txt', displaying several lines of structured text.

**Tornado**

**CONSTITUTION DU CONVERTISSEUR**

**Légende élément**

Profils logiciels

STRUCTURE

PROFILS HYBRIDES

1 prétraitements sur profils externes

séparation des fichiers :

1 un par apprenant

2 identification du type de fichiers

3 description de la structure des fichiers

3 EXÉCUTION

Description des données - Etape 1 sur 2

Informations sur le Tourbillon

Séparation des fichiers

Description des données

Création du Tourbillon

Vérification du Tourbillon

Exécution du Tourbillon

Type de données

- liste d'éléments (fichier texte contenant un élément par ligne)
- fichier avec séparateurs (un ou plusieurs caractères donnés séparent chaque champ)
- fichier avec largeur fixe (les champs sont alignés en colonnes)
- fichier XML
- fichier avec balises ouvrante/fermante (les éléments sont entre deux balises)
- base de données

Aperçu du fichierformid\_traces\_LM1\_2005\_03\_24\_18\_12\_52\_T1.txt

```
TLM1#mustild#1#C#c
TLM1#mustild#3#V#1#E#u=30.0;r=2.0;j=15.0;fuse=50;
TLM1#mustild#3#K#1#1#2#u=47.0;r=2.0;j=23.5;fuse=50;
TLM1#mustild#3#K#1#1#2#u=50.0;r=3.0;j=16.666666666666668;fuse=50;
TLM1#mustild#3#V#1#S#u=80.0;r=31.0;i=2.5806451612903225;fuse=50;
TLM1#mustild#0#C#d
```

Annuler << Précédent Suivant >> Terminer ? Aide

# Prétraitement des profils :

## Description des données

**Tornado**

**CONSTITUTION DU CONVERTISSEUR**

**Légende élément**

1 **prétraitements sur profils externes**

séparation des fichiers :  
1 un par apprenant

2 identification du type de fichiers

3 description de la structure des fichiers

Profils logiciels

STRUCTURE

PROFILS HYBRIDES

EXÉCUTION

**Description des données - Etape 2 sur 2**

Informations sur le Tourbillon | Séparation des fichiers | **Description des données** | Création du Tourbillon | Vérification du Tourbillon | Exécution du Tourbillon

Cette étape vous permet de choisir le séparateur contenu dans vos données.

Dernière Foi : Espace % :

Séparateur20

Tabulation (carré)     Égal     Point-virgule

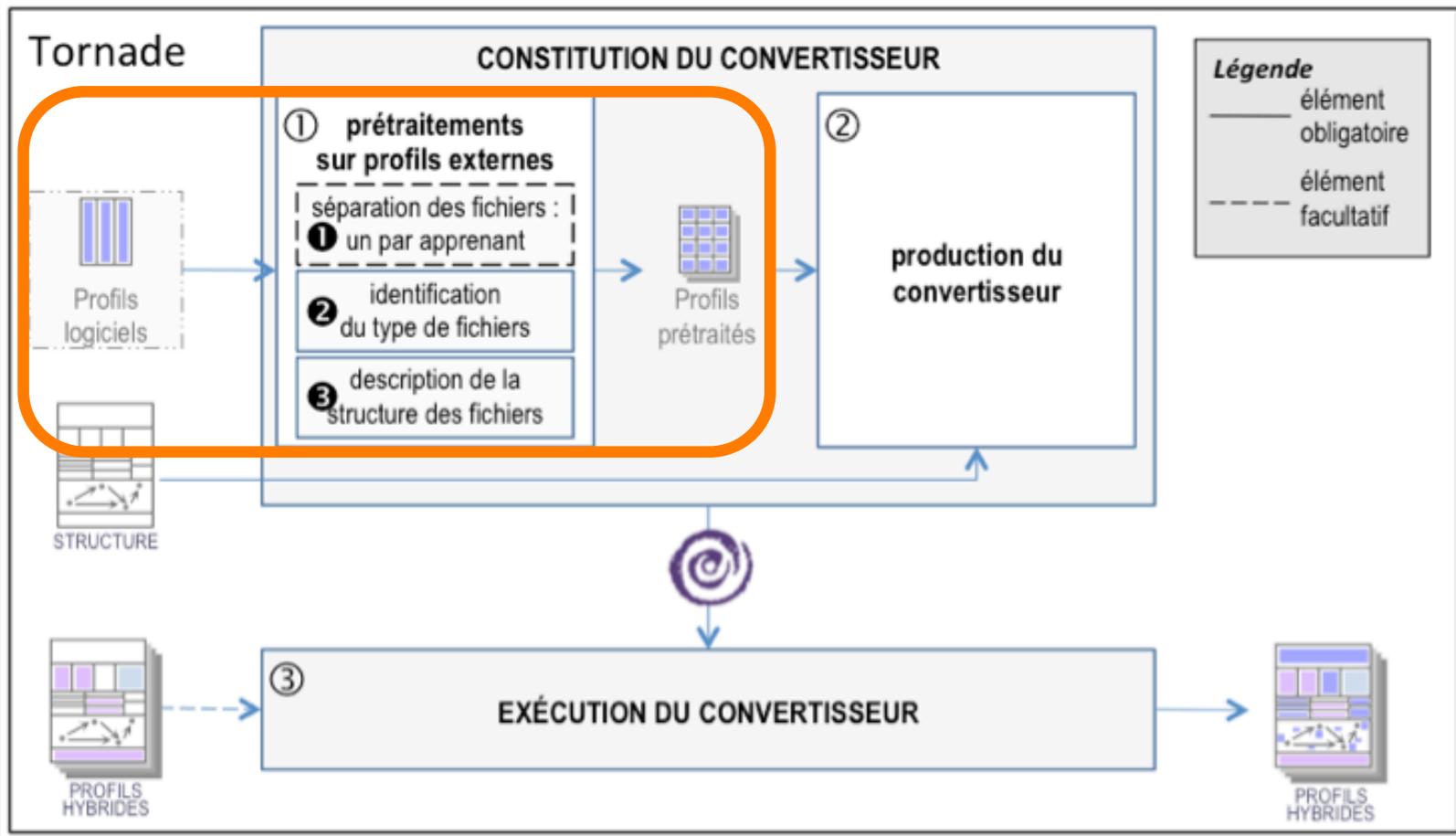
Espace     :=     Autre :

Aperçu des données

-	1	Cours		les	différentes	poubelles
-	Conteneur	à	verre			
-	Poubelle	à	recyclage			
-	Bac	à	compostage			
-	Poubelle	normale				
-	Fin					
-	2	Cours		recyclage		
-	Recyclage	du	verre			
-	Recyclage	du	papier			
-	Recyclage	du	plastique			
-	Recyclage	du	métal			

**X Annuler**    << Précédent    Suivant >>    Terminer    ? Aide

# Obtention de profils prétraités



# Production d'un convertisseur

## Tornado

### CONSTITUTION DU CONVERTISSEUR

Création du Tourbillon

Informations sur le Tourbillon | Séparation des fichiers | Description des données | **Création du Tourbillon** | Vérification du Tourbillon | Exécution du Tourbillon

#### Données du logiciel externe

07	-	3	Jeu			correspondance	dechets-poubelles
07	-	Niveau	1				
15	-	Niveau	2				
17	-	Niveau	1				
48	-	Diagnostic					
coups réussis	réussis		40	2,	nombre	total	de coups

#### Parcours des briques de la structure EcoleEte

##### Parcours des éléments de la brique Tri selectif de type Liste

Information sur la brique :

Date de l'évaluation : 13/10/2010  Sélection dans le fichier externe

Évaluation source :

Commentaire :   Juxtaposition

Information sur lescomposantes :  Moyenne  Somme

composante : Correspondances dechets-poubelles	
valeur (échelle numérique)	40
composante : Correspondances couleurs-poubelles	
valeur (échelle numérique)	
composante : Chercher l'intrus	

**production du convertisseur**

**SEUR**

PROFILS HYBRIDES

**Légende**

— élément obligatoire

- - - élément facultatif

13/12/12

TICE 2012 - Lyon

( 17 )

( 24 )

# Vérification du convertisseur

**Tornado**

**CONSTITUTION DU CONVERTISSEUR**

**Légende**

- élément obligatoire
- - - élément facultatif

Confirmation ou modification du Tourbillon

Informations sur le Tourbillon   Séparation des fichiers   Description des données   Création du Tourbillon   **Vérification du Tourbillon**   Exécution du Tourbillon

Tourbillon mis en route sur le fichier : Hugo\_profil.txt

il y a 1 erreur(s) sur le profil créé, elle(s) est (sont) caractérisée(s) par des cases rouges

sous_composante	nom	Verbes du 2eme groupe			
valeur					
sous_composante	nom	Verbes du 3eme groupe			
valeur					
brique	id	4	type	3	date
texte					
brique	id	8	type	0	nom
infos_echelle					
echelle	id	0			
arbre_des_composantes	niveaux	2	ponderation	faux	
composante	nom	Addition			
valeur		Connaitre ses tables			
sous_composante	nom	Savoir résoudre des problèmes			
valeur		nomabrege	Savoir résoudre problemes		
composante	nom	Multiplication			
valeur		commentaire			
brique	id	10	type	0	nom
infos_echelle					
echelle	id	17			
arbre_des_composantes	niveaux	1	ponderation	faux	
composante	nom	Correspondances dechets-poubelles			
valeur	valeur	1	nomabrege	Dechets-Poubelles	
composante	nom	Correspondances couleurs-poubelles			
valeur	valeur		nomabrege	Couleurs-Poubelles	
composante	nom	ERREUR : selection hors echelle			
valeur		Chercher l'intrus			
composante	nom	Recyclage			
valeur					
composante	nom	La chanson des poubelles			
valeur					

**production du convertisseur**

**SEUR**

PROFILS HYBRIDES

**X** Annuler

Corriger

**✓** Confirmer

Terminer

? Aide

13/12/12

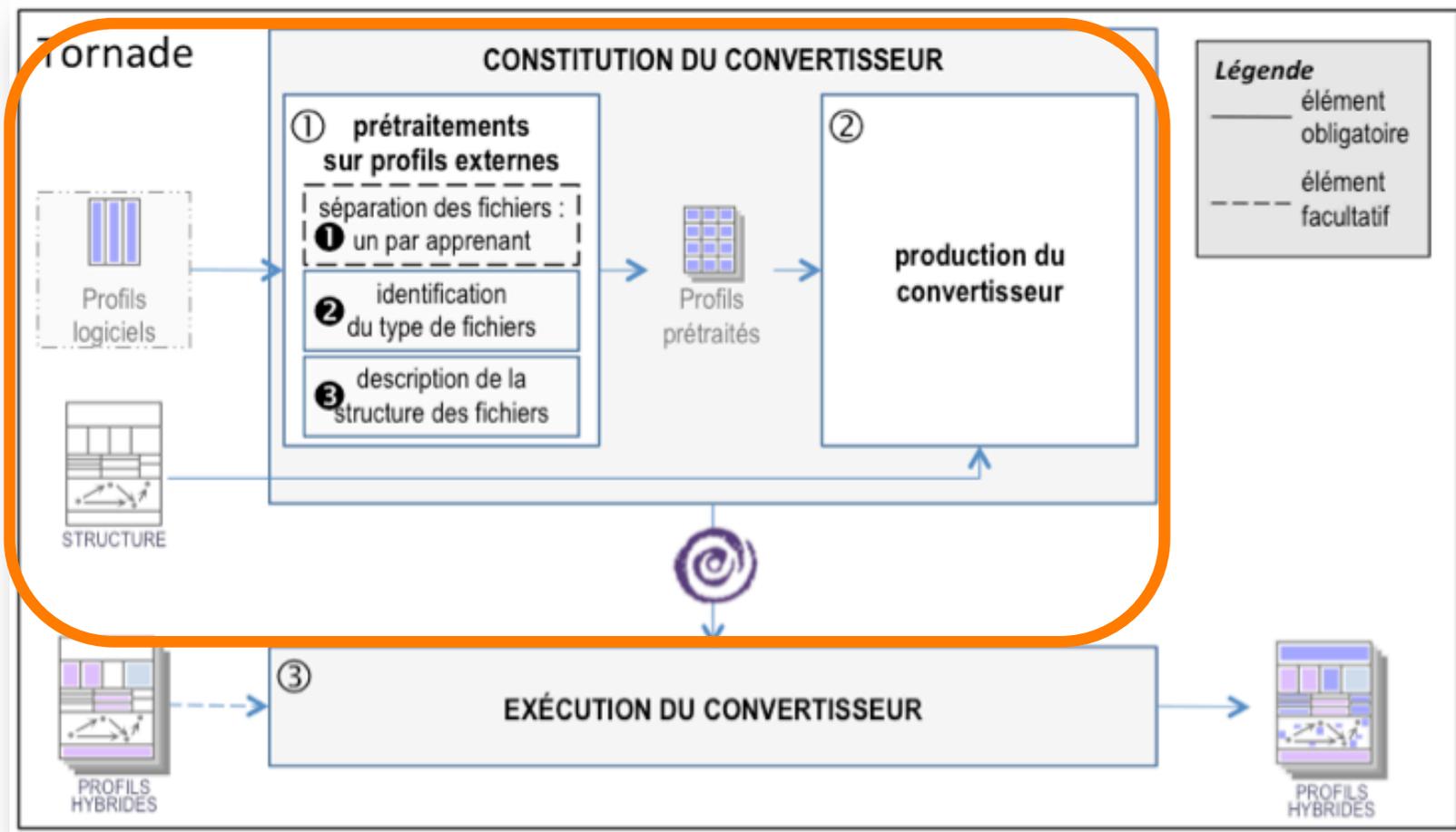
TICE 2012 - Lyon

18

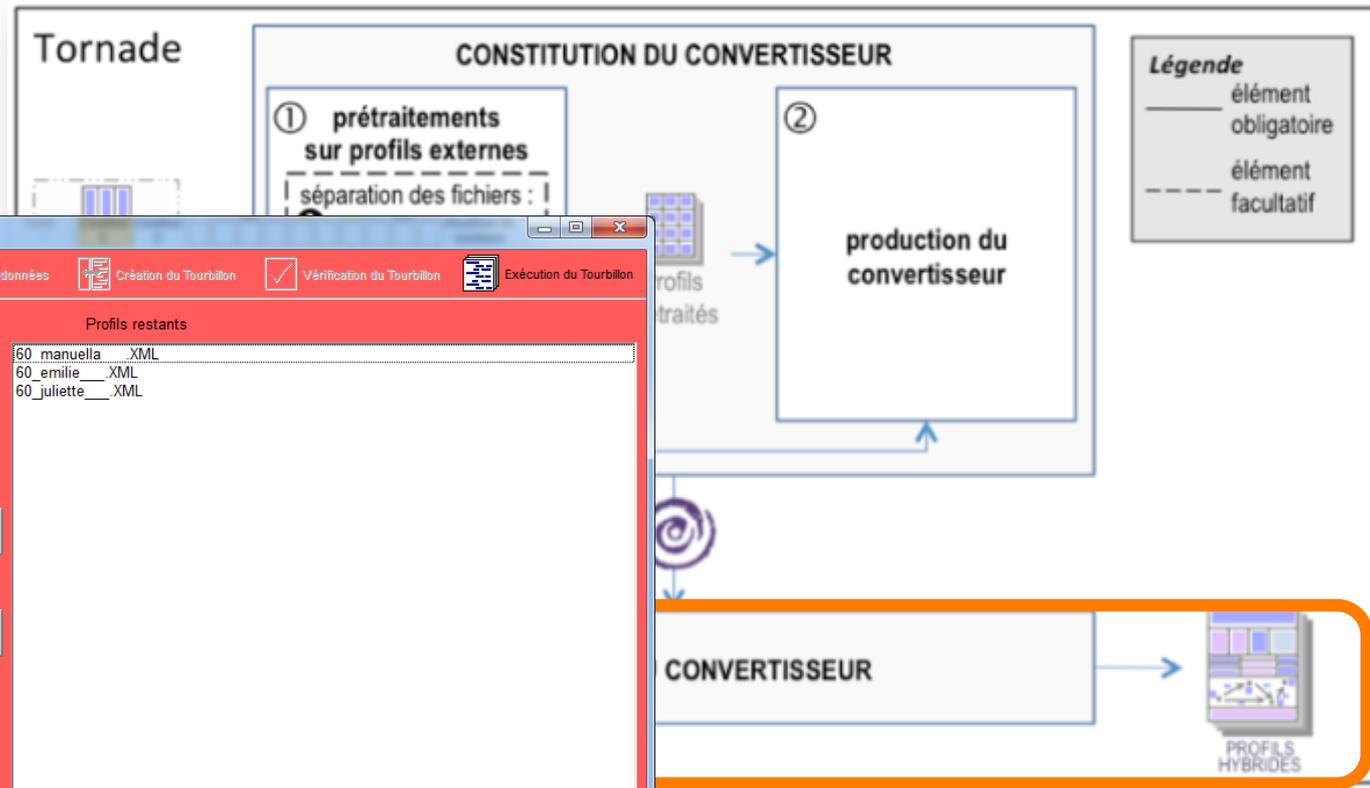
---

24

# Obtention d'un convertisseur pour une structure de profils



# Exécution d'un convertisseur



Exécution du Tourbillon

Informations sur le Tourbillon Séparation des fichiers Description des données Création du Tourbillon Vérification du Tourbillon Exécution du Tourbillon

Élèves	Profils	Profils restants
Alain		60_manuella__XML
Benoit		60_emilie__XML
Chloe		60_juliette__XML
Damien		
Emma		
Florian		
Guillaume		
Hugo	60_hugo__XML	
Isabelle		
Julien		
Kévin	60_kevin__XML	
Laetitia		
Mona		
Noemie		
Ophelie		
Pierre		
Quentin		
Robert		
Sabrina		
Tristan		
Ulysse		
Veronique		
...		

Ajouter un élève... Convertir un autre fichier...

Annuler << Précédent Suivant >> Terminer Aide

# Résultat de la conversion

- 1 structure de profil

```
date_derniere_modif="11/01/2010" nom_eleve="" prenom_eleve="">
<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">
  <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>
```

- n profils complétés

```
<structure id="0" nom="Echelle" createur="marc" 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 sélectif" 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">
    <evaluation date="2011/02/01" source="">60 </evaluation>
  </valeur>
</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>
```

```
<evaluation date="2011/02/01" source="">100</evaluation>
</valeur>
</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>
```

# Dans Tornade, qui fait quoi ?

- Phase de prétraitement des données



- faite une seule fois par EIAH
- par un enseignant-expert
- nécessite des connaissances sur les fichiers

- Phase de production du convertisseur



- faite une seule fois pour un couple EIAH / structure de profils
- par un enseignant-expert
- nécessite des connaissances sur les profils et l'EIAH

- Phase d'exécution du convertisseur



- faite chaque fois qu'on veut créer ou mettre à jour des profils
- par un enseignant
- nécessite des connaissances sur les profils, les élèves

# Évaluation

- Utilité
  - test sur un trentaine de profils issus de logiciels éducatifs variés
  - difficultés rencontrées
    - formats de fichiers incompatibles
    - répartition des données non gérées
      - ex : trouver les n dernières occurrences d'un élément
    - des compléments nécessaires
- Utilisabilité
  - mise à l'essai
    - 2 enseignants, 60 étudiants
    - conversion des données réussies
  - mise en évidence la complexité du logiciel
    - partie création d'un convertisseur
  - étude ergonomique afin de simplifier l'interface
    - refonte du module nécessaire

# Conclusion

- Problématique
  - réutilisation de profils d'apprenants issus de sources hétérogènes dont le formalisme de représentation n'est pas connu
- Solution proposée
  - démarche générique
  - définition d'un processus d'intégration des données
  - mise en œuvre dans le module Tornade de l'environnement EPROFILEA
- Exploitation des résultats
  - combiné aux autres modules d'EPROFILEA, Tornade permet d'obtenir des **profils hybrides évolutifs**
    - données issues de logiciels et de sources papier
    - données mises à jour en fonction de l'évolution des apprenants
  - d'autres modules permettent maintenant de proposer
    - des **activités sur les profils** (visualisations riches, profil cible...)
    - des **activités personnalisées en fonction du contenu des profils**
      - sur papier
      - au sein de logiciels éducatifs existants

