



HAL
open science

**Concevoir, produire, décrire, évaluer, améliorer,
partager et conserver des données, opérateurs, processus
d'analyse et résultats d'études sur les logs d'activités
d'apprentissages humains avec ordinateur**

Denis Bouhineau, Vanda Luengo, Nadine Mandran

► **To cite this version:**

Denis Bouhineau, Vanda Luengo, Nadine Mandran. Concevoir, produire, décrire, évaluer, améliorer, partager et conserver des données, opérateurs, processus d'analyse et résultats d'études sur les logs d'activités d'apprentissages humains avec ordinateur. 2018. hal-01843970

HAL Id: hal-01843970

<https://hal.science/hal-01843970>

Preprint submitted on 19 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Concevoir, produire, décrire, évaluer, améliorer, partager et conserver des données, opérateurs, processus d'analyse et résultats d'études sur les logs d'activités d'apprentissages humains avec ordinateur

► Denis BOUHINEAU (LIG, Université Grenoble Alpes),
Vanda LUENGO (LIP6, Sorbonne Universités, UPMC Université Paris 06),
Nadine MANDRAN (LIG, Université Grenoble Alpes).

■ **RÉSUMÉ** • Ayant fait le constat que les validations expérimentales de la recherche en EIAH sont particulièrement coûteuses, nous proposons une architecture et un ensemble cohérent de services permettant la structuration d'une plateforme web de conception et d'échange de données, d'opérateurs, de processus d'analyse et de résultats intervenant dans les études d'interactions homme-machine lors d'apprentissage avec ordinateur. Notre proposition met en avant les cycles de vie des artefacts en jeu et se place dans la perspective d'une plateforme de soutien effectif à la recherche en cours de réalisation permettant un travail collaboratif entre chercheurs et un rapprochement bénéfique données-analyses. Un prototype fonctionnel de la plateforme visée, précurseur de ladite proposition est décrit avec les premiers indicateurs d'utilisation et les premiers résultats.

■ **MOTS-CLÉS** • Plateforme, pluridisciplinarité, capitalisation, analyse, trace.

■ **ABSTRACT** • *Experimental validations of TEL research are particularly costly, we propose an architecture and a set of services allowing the structuring of a platform for design and exchange of data, operators, analysis processes and results for man-machine interaction studies in TEL. Our proposal is placed in the perspective of a platform for support to research in progress allowing collaboration and a beneficial combination of data and analyses. A functional prototype is described; first indicators of use and results obtained are discussed.*

■ **KEYWORDS** • *Platform, Multidisciplinary, Capitalization, Analysis, Traces.*

1. Introduction

Depuis les années 1990-2000, la création de l'acronyme EIAH (Environnements Informatiques pour l'Apprentissage Humain) puis la réunion en 2003 (Desmoulin *et al.*, 2003) des conférences EIAO, Environnements interactifs d'apprentissage avec ordinateur (Desmoulin *et al.*, 2001) et Hypermédias et Apprentissages (de Vries *et al.*, 2001), une communauté large s'est constituée regroupant enseignants et chercheurs de nombreuses disciplines distinctes : sciences de l'éducation, didactique, psychologie, informatique et sciences des données. Les résultats visés par cette communauté nécessitent des apports simultanés et cohérents venant non seulement de ces différentes disciplines, mais de surcroît produits par des individus ayant des métiers différents qu'il s'agisse de praticiens (enseignants, décideurs), de chercheurs de l'apprentissage (chercheurs en sciences de l'éducation, didacticiens, psychologues, sociologues) ou des chercheurs et techniciens de l'information (informaticiens, statisticiens). L'exemple abordé dans cette communication concernera donc la production et l'analyse de traces d'apprentissage permettant la validation de la recherche EIAH où :

- les praticiens sont nécessaires pour mobiliser les apprenants fournissant ces traces,
- les chercheurs de l'apprentissage sont nécessaires pour concevoir les protocoles expérimentaux et la modélisation de ces traces,
- les chercheurs et techniciens de l'information sont nécessaires pour récolter ces traces, concevoir et mettre en œuvre les processus d'analyse.

L'ensemble des disciplines, sciences de l'éducation, didactique, psychologie, informatique et sciences des données, est mobilisé.

Concernant la validation ou l'évaluation des résultats de la recherche en EIAH, du point de vue informatique le domaine des EIAH fait appel à plusieurs disciplines (IA, IC, SI, IHM) et au sein de ces domaines les méthodes d'évaluation ne sont pas les mêmes. Il est de même en ce qui concerne les domaines SHS. En effet, il est possible qu'une recherche en didactique soit uniquement qualitative en même temps qu'en psychologie de l'apprentissage elle soit uniquement quantitative. L'invariant dans ces recherches est celui d'un ou plusieurs utilisateurs en interaction avec un système informatique ou médié par un système informatique. Un postulat ici est donc que ces interactions puissent être tracées. Enfin, la recherche en EIAH est une recherche où par définition l'humain est au centre. Aussi, il est donc difficile de proposer des méthodes et des modèles de validation uniques

ayant une forme de preuve formelle. Lors de cet article nous traiterons donc également de la question de l'évaluation des recherches EIAH en étudiant comment une démarche centrée sur les traces peut permettre d'améliorer ce processus de validation et nos recherches dans le domaine des EIAH.

Enfin, ainsi formulé, il apparaît clairement que l'obtention de résultats de validation en production et analyse de traces d'apprentissage est fort coûteuse. C'est une expérience bien connue des chercheurs et doctorants du domaine. Une solution pour réduire ce coût est de mettre en place des outils de travail collaboratif pour supporter la conception, la production et l'évaluation des données, des algorithmes, des processus d'analyses et des résultats intervenant dans les études en EIAH.

Plusieurs projets et plateformes depuis plus de 10 ans ont déjà eu l'objectif de partager et capitaliser les éléments d'une recherche en EIAH. L'objectif visait toutefois rarement l'ensemble du panorama et s'attachait plus à archiver le travail effectué qu'à fournir les moyens de l'effectuer :

- Souvent, les données récoltées, fichiers logs, chats, etc. étaient au centre des efforts ; par extension les motivations de la recherche, le contexte d'acquisition des données, les processus d'analyse, les résultats obtenus étaient abordés, mais dans une moindre mesure ou plus tardivement, cf. le projet DataShop (Koedinger *et al.*, 2010).

- L'un des cadres de référence évoqués comme support à ces projets pouvait être celui des archives ouvertes « Open Archive Initiative » (Nelson et Warner, 2002) situant le projet au-delà du contexte de l'activité en cours, dans le contexte d'une conservation d'une activité arrivée éventuellement à son terme, pour le temps long, cf. le projet Mulce (Reffay *et al.*, 2012).

Ces deux projets, DataShop (Koedinger *et al.*, 2010) et Mulce (Reffay *et al.*, 2012), en particuliers, ont en commun d'avoir eu une existence suffisamment longue pour pouvoir tirer quelques leçons. Tout d'abord que l'entreprise générale est viable et peut fournir à terme le support visé (c'est-à-dire, en prenant en compte les objectifs initiaux de chaque projet en question). Ensuite, que la réalisation d'une telle ambition n'est pas sans difficulté et nécessite de nombreuses décisions. En particulier ces deux projets, considérant les données brutes à produire et à conserver, ont fait le constat qu'il se trouve là, des questions pratiques et techniques inévitables, menant à des choix parfois délicats et ayant des répercussions importantes sur l'ensemble du projet. Pour ces questions et les choix associés (quel(s)

format(s) utiliser pour les données, comment prendre en compte des données aux formats différents ?) DataShop et Mulce ont eu des réponses différentes.

Pour dépasser le problème de la multiplicité des formats, DataShop a imposé un ensemble de 3 standards pour partager (1) les données d'interaction entre apprenants et EIAH, (2) les modèles d'apprenant et (3) les résultats d'analyse. Ces 3 formats sont liés, entre autres, à l'utilisation des tuteurs intelligents développés au Pittsburgh Science of Learning Center. Les formats sont techniquement définis à un niveau très bas de description mais permettent de véhiculer des informations particulièrement riches du point de vue sémantique vis-à-vis de l'apprentissage et du contexte (description des problèmes en suite d'étapes reliées à des éléments de connaissance). Sans être des formats « propriétaires / fermés », ils imposent cependant une rigidité dans la description des traces et nécessitent un modèle de connaissance avancé pour la modélisation des situations.

Face à ce choix, Mulce privilégie la flexibilité pour prendre en compte des recherches en EIAH venant d'horizons multiples et a choisi de ne pas être lié à un format particulier ou contraignant ou à des logiciels particuliers pour la conservation des données : les moyens d'archivages utilisés doivent être structurés mais ouverts, interopérables et agnostiques des contextes de production. En contrepartie, pour pallier à la technicité de certaines solutions employées, l'archive d'une étude dans Mulce doit être associée à des descriptions textuelles, lisibles par les chercheurs afin de permettre une bonne compréhension des données au niveau sémantique et rendre ainsi ces documents humainement auto-descriptifs et pleinement utilisables par les chercheurs. L'archive technique est doublée par une description humaine.

La question du format des données est un exemple particulièrement frappant parmi ceux qui orientent profondément un projet de plateforme pour la recherche en EIAH. Parmi les questions et choix définissant une plateforme, il se place à un niveau technique et pratique souvent de bas niveau, proche des données. Une question similaire peut exister pour le format des analyses : Quel format choisir pour concevoir, partager, conserver une analyse ? Nous pouvons énumérer quelques possibilités :

- une forme textuelle informelle pour une description narrative ;
- un format textuel plus rigoureux avec la liste des articles publiés concernant l'analyse et la description des outils mathématiques ou algorithmiques employés ;

- un format plus proche de l'objet technique comportant une archive des scripts et programmes, dans un langage de programmation particulier, effectivement utilisés ;
- un format ontologique décrivant les relations sémantiques entre les différents composants ;
- ou encore un format ad hoc associant les uns et les autres ?

De même, quel format choisir pour reporter les résultats entre ceux favorisant les objets techniques obtenus, ceux concernant les visualisations produites, ceux pour les commentaires associés et enfin ceux des publications officielles ?

Aux niveaux plus élevés, moins techniques et plus fonctionnels (organisationnels ou stratégiques), se trouvent d'autres choix directeurs pouvant guider la réalisation d'une telle plateforme. Un exemple de choix directeur, quelle est l'activité envisagée pour les utilisateurs de la plateforme :

- une activité centrée sur l'exploration intellectuelle (préparation de projets) ou la recherche effective (exploratoire) ;
- une activité de diffusion des objets obtenus après leur validation (vitrine ou catalogue) ;
- une activité de travail collaboratif (entre chercheurs partageant un même projet de recherche) ou de suivi (encadrement de thèse) ?

Un autre exemple de choix directeur, quelle communauté est visée : une communauté scientifique réduite (EIAH), plus large (Recherche en Informatique Centrée Humain), encore plus large (Data Science) ? Et parmi ces communautés, quels utilisateurs : les jeunes chercheurs (en thèse, pour un guidage et un suivi) ; les praticiens à la recherche d'objets techniques concrets (plus ou moins prêts à l'emploi) ; les décideurs (en recherche de visualisations directement compréhensibles, de tableaux de bord exploitables et de descriptions factuelles simples) ? Ces décisions sont celles qui sont nécessaires pour définir le type de plateforme de soutien à la recherche expérimentale en EIAH visé.

Initialement, DataShop¹ et Mulce ont effectué des choix orientés données, archivage, tuteurs intelligents, apprentissage en ligne, etc. Pour notre part, nos choix se sont portés sur un panorama englobant données, opérateurs et processus d'analyse pour des activités collaboratives en cours de réalisation

¹ DataShop a progressivement évolué en proposant des algorithmes (performance, difficulté, courbes d'apprentissage) et une gestion de processus d'analyse (avec un format fixé associé).

de recherche en EIAH basées sur l'analyse des logs d'interactions entre apprenants et EIAH (sans contraintes sur le type d'EIAH).

Notre communication concerne ainsi une investigation des éléments clés (artefacts pour l'architecture et services nécessaires aux activités visées) autorisant une décomposition de la recherche expérimentale en EIAH pour élaborer le modèle d'une plateforme support à une activité de recherche permettant un travail collaboratif (section 3, notre proposition essentielle). Cette investigation s'est effectuée en prenant appui sur des analyses des cycles de vie de la recherche en EIAH observés dans la littérature (section 2) et des retours d'expérience d'un projet (section 4) qui ont servi à la fois d'initiateur à cette recherche et qui servent aujourd'hui de prototype pour valider une partie de notre proposition. La suite de la communication comporte donc 3 parties :

- Analyse des cycles de vie des données et opérateurs algorithmiques ou statistiques utilisés lors des études expérimentales d'interactions apprenants-machine en EIAH
- Modèles pour une architecture de plateforme de soutien à la recherche expérimentale en EIAH et pour un ensemble de services cohérents associés
- Descriptions d'un prototype fonctionnel précurseur de notre proposition et outil support de notre modèle.

2. Cycles de vie des études expérimentales d'interactions apprenants-machine en EIAH

L'analyse des activités de recherche expérimentale en EIAH a été effectuée en particulier pour ce qui concerne le cycle de vie des données. Ainsi, (Romero et Ventura, 2007), (Stamper *et al.*, 2011) ont dégagé différentes phases : (1) préparation ou conception abstraite d'un modèle de données en amont de l'expérimentation et de la récupération des données, (2) production ou collecte des données, (3) vérification des données obtenues et validation de leur qualité, (4) enrichissement et analyse en vue de répondre à une question de recherche.

À cette organisation en phases successives s'ajoutent deux niveaux de boucles, représentatifs de la complexité de la recherche, introduisant deux types de cycle. Le premier niveau concerne un cycle global comprenant l'ensemble des phases de la vie des données (1)+(2)+(3)+(4). Il concerne le travail qui recommence quand les analyses en fin d'activité ont produit de

nouvelles questions de recherche menant à une nouvelle expérimentation, ou à une reprise d'une partie de l'expérimentation pour obtenir des résultats complémentaires. Après (4) il peut y avoir un retour à (1), (2) ou (3). Le second niveau concerne chaque phase de l'activité et est lié au cycle de vie de la phase, menant à d'éventuelles répétitions de la phase ou à des retours sur les phases précédentes. Après une phase (i), cette même phase peut être relancée, ou une phase précédente (i-1), (i-2). L'existence de ces cycles et retours arrières est peu abordée, formalisée, exploitée dans l'analyse de l'activité dans les recherches expérimentales en EIAH, pour autant elle constitue une part importante de l'activité et nous la prendrons en compte dans notre modèle de plateforme de soutien à la recherche expérimentale en EIAH.

Concernant les éléments produits lors de ces recherches, une part des activités concerne des moments d'organisation donnant lieu à des rédactions de documents descriptifs ou narratifs (pour les objectifs visés, les moyens employés, les contextes rencontrés, etc.) et une part concerne des productions plus techniques relatives aux traces convoitées côté « données » (avec en particulier, les jeux de données eux-mêmes) et aux scripts utilisés, algorithmes produits, etc. pour le côté « algorithmique ». Les analyses effectuées dans (Romero et Ventura, 2007) et (Stamper *et al.*, 2011) montrent que le plus souvent il existe un focus important sur les éléments techniques produits se concentrant essentiellement sur les données, oubliant d'une part la production informelle et le côté algorithmique. Contexte, organisation, etc. sont souvent limités à des métadonnées formatées et réduites associées aux données ce qui ne permet pas de retracer l'ensemble de l'activité. Si les documents informels ont parfois été pris en compte dans certains projets (par exemple (Reffay *et al.*, 2012) dans le projet Mulce), c'est un cas assez rare pour être signalé. Notre proposition pour une plateforme de soutien à la recherche suivra sur ce point ce projet.

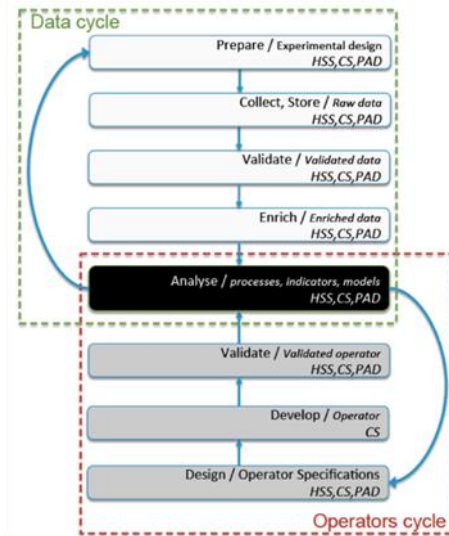


Figure 1. DOP 8. Cycle de vie des données et des opérateurs dans les études expérimentales en EIAH d’après (Mandran *et al.*, 2015). En particulier, à observer, la similarité des deux cycles et leur point de convergence lors de l’exécution effective de l’analyse proprement dite sur les données définies pour une étude.

Dans (Mandran *et al.*, 2015), voir figure 1, la partie algorithmique a été prise en compte considérant la vie des données en relation avec la vie des opérateurs algorithmiques ou statistiques intervenant dans les processus d’analyses. L’analyse produite par (Mandran *et al.*, 2015) décrit la recherche en se focalisant sur les artefacts techniques « données » et « opérateurs » intervenant en EIAH et propose une modélisation de l’activité selon un double cycle en 8 dont le centre ou point de convergence des deux cycles est constitué par l’exécution effective de l’analyse proprement dite sur les données. Ce double cycle en 8 montre en particulier la similarité des deux cycles « données », « opérateurs » avec dans les deux cas (1) une phase préparatoire, (2) une phase de production des artefacts techniques, (3) une phase de validation de ces artefacts et (4) une phase d’utilisation de ces artefacts où données et opérateurs interviennent en un point de rencontre des deux cycles. L’introduction d’un double cycle de vie permet de dissocier l’activité de recherche en deux avec d’un côté une ingénierie des données et de l’autre une ingénierie des algorithmes. Cette décomposition favorise un découpage de l’activité de recherche en deux parties plus simples qui peuvent être réalisées indépendamment, éventuellement par des individus

distincts, ayant des métiers différents, travaillant en collaboration. Dans l'hypothèse où données et opérateurs se trouvent sur une même plateforme, un spécialiste des données pourra trouver un opérateur algorithmique ou statistique sur la plateforme pour analyser ses données sans avoir à le produire et se soucier de son cycle de vie et le concepteur d'un opérateur pourra tester son exécution sur des données de la plateforme sans avoir à les produire et se soucier de leur cycle de vie. La similarité des cycles de vie doit permettre au niveau de la conception et de la mise en place de la plateforme d'avoir des gains de productivité dans la réalisation des éléments similaires et au niveau des usages, elle doit permettre aux utilisateurs d'utiliser les parties qui les concernent moins avec une plus grande aisance.

L'analyse menant au double cycle en 8 (Mandran *et al.*, 2015) met en lien le cycle de vie des données et le cycle de vie des opérateurs, c'est une extension des analyses faites autour de DataShop (Stamper *et al.*, 2011) et de l'état de l'art (Romero et Ventura, 2007), nous proposons de la prolonger pour bénéficier des avantages associés (pour le développement et les usages) en : (1) explicitant un cycle de vie des processus d'analyse, (2) introduisant les résultats obtenus (indicateurs, modélisation, etc.) comme artefacts ayant une existence et un cycle de vie à prendre en compte. Ainsi l'activité de recherche expérimentale en EIAH sera décrite selon un ensemble de 4 artefacts, chacun plus simples que l'ensemble, et 4 cycles de vie similaires, en lien les uns avec les autres, comprenant :

- « données » (cf. préc.),
- « opérateurs » (cf. préc.),
- « processus d'analyse » (objet, à l'intersection de l'univers des données et de l'univers algorithmique, au centre du double cycle en 8, issu de l'association des opérateurs sous forme d'un processus complet pour être exécutés sur des données) et
- « résultats » (produit du processus d'analyse décrit à la ligne précédente).

C'est le cœur de notre proposition pour la définition d'une architecture et d'un ensemble cohérent de services permettant la structuration d'une plateforme de soutien à la recherche expérimentales en EIAH. La proposition complète constitue l'objet de la section suivante.

3. Modèles pour une architecture et un ensemble de services pour une plateforme de soutien à la recherche expérimentale en EIAH

L'analyse précédente a mis en évidence l'existence de 4 artefacts techniques intervenant dans une recherche expérimentale en EIAH : « données », « opérateurs », « processus d'analyse » et « résultats » associés à des objets informatiques spécifiques (fichiers de traces d'interaction entre un apprenant et un EIAH, fichiers sources des programmes et scripts des opérateurs et des processus d'analyse, fichier mixte de traces, de sources de programmes, de modèles et d'indicateurs pouvant être présentés sous forme visuelle au travers de tableaux de bord ou d'outils spécifiques pour les résultats). Pour structurer la plateforme il nous semble important d'ajouter à ces 4 artefacts, un cinquième, global, représentant l'ensemble de l'activité, nous le nommerons « étude ». Faisons un tour d'horizon des cinq éléments architecturaux obtenus et des outils de gestion associés.

3.1. Ensemble de 5 artefacts pour une plateforme EIAH

3.1.1. Études

La notion « étude » a pour rôle d'assurer une vision globale de l'activité du chercheur ou du groupe de chercheurs impliqué et faire le lien entre données, analyses et résultats. Elle doit permettre la prise en compte d'une recherche ponctuelle ou d'un ensemble de recherches coordonnées. Dans un premier temps, elle rassemble les métadonnées de l'activité à un niveau global, son contexte. Mais outre les métadonnées globales, quels autres artefacts lui sont associés ? Une activité de recherche est définie par un projet faisant intervenir un ensemble de personnes ayant un ensemble d'objectifs scientifiques à réaliser dans un contexte spécifique et un temps particulier. La mise en place de ce projet, son déroulement dans le temps sont associées, certes aux « données », « processus d'analyse » et « résultats », mais également à l'ensemble des documents informels échangés parmi les participants à l'étude, ou relatant l'histoire et l'avancement de la recherche, les a priori théoriques, épistémologiques ou didactiques, le contexte, etc.

Les artefacts liés à une « étude » sont donc constitués d'une partie formelle et technique comprenant des métadonnées (personnes, dates, sujets, domaine, licences, etc.), et d'une partie informelle prenant en compte toute la vie de l'activité (comportant documents, descriptions, permettant échanges, discussions pratiques ou éthiques, etc.).

En pratique, pour prendre en compte la partie informelle nous proposons pour la partie « étude » l'adjonction d'utilitaires de gestion électronique de documents (GED), de zone de description ou de prises de notes informelles, suffisamment riches et d'outils d'échanges sociaux (forum, chat, etc.)

3.1.2. Données

Les « données » peuvent comporter des traces d'interaction enseignant ou apprenant avec un EIAH, des données qualitatives, etc. Elles sont associées à des métadonnées et éléments de description du contexte (en particulier les conditions de production de ces données, la description des éléments observés, etc.)

Pour les « données », en tant qu'artefact, les solutions logicielles les plus évidentes pour les prendre en compte dans une architecture sont évidemment les systèmes de gestion de bases de données (SGBD, qu'elles soient relationnelles, décisionnelles, NoSql, etc.) avec des interfaces vers/venant de l'univers des fichiers (csv ou autres).

La structure des données pouvant être assez immédiate (souvent « simples » log et métadonnées), quelques une des attentes particulières du modèle de plateforme EIAH visé peuvent être prises en compte :

Outre les fonctionnalités usuelles attendues d'un SGBD (création d'ensemble de données, ajout de données à ces ensembles, conservation de ces données, gestion de droits d'accès à ces données, analyses, interrogations et visualisations simples de ces données) plusieurs fonctionnalités spécifiques sont également à considérer : la capacité à pouvoir gérer des volumes de données importants (imaginez l'ensemble des réponses aux questionnaires disponibles sur un serveur de Mooc ayant un peu de popularité et un peu d'ancienneté ...), la capacité à gérer l'histoire des jeux de données, éventuellement leurs différentes versions (sachant que ces données sont essentiellement obtenues par accumulation successives avec peu de modification). À l'intersection de ces deux attentes se trouvent par exemple des gestionnaires de base de données NoSql tels qu'Exist (Siegel & Retter, 2014), CouchDB (Anderson *et al.*, 2010).

3.1.3. Processus d'analyse

Le « processus d'analyse » est constitué de l'association de données et d'opérateurs en suites de tâches à exécuter (workflow en anglais). Il est lui associé à des métadonnées et éléments de description du contexte.

Pour les « processus d'analyse », divers outils ont été mis en place dans les communautés d'analyse de données (Demsar *et al.*, 2013), (Bellanger et Tomassone, 2014), (Frank *et al.*, 2016) ainsi que chez de nombreuses entreprises (SAS, SPSS) ou dans le monde des logiciels libres (R, RapidMiner, orange). Plusieurs tendances sont notables : approches statistiques, fouilles de données (informatique, IA), approches centrées humains (décisionnelles, analytics). À notre connaissance, aucune suite logicielle d'analyse n'a été conçue spécifiquement pour le domaine de l'EIAH. L'ouverture et la capacité d'extension de ces solutions logicielles sont un critère de choix pour déterminer quelle solution adopter dans une architecture pour l'EIAH ; ainsi, la plupart des solutions logicielles ouvertes sont à privilégier. Cependant, ces solutions logicielles ne prennent en charge que la partie algorithmique de l'activité, pour ce qui concerne les autres étapes du cycle de vie des analyses (conception, validation, conservation, documentation, partage, retour arrière, etc.), étant donnée la nature algorithmique des processus d'analyse, nous emploierons une solution logicielle de type forge logicielle² dans l'architecture pour gérer les spécificités de ces activités, en particulier. Nous pourrions ainsi profiter de la gestion des aspects informels liés au développement logiciel (partage, documentation, forum d'échange, gestion d'anomalies, versionning, etc.). Car pour la prise en compte de l'informel et des boucles de retour arrière, les communautés de développement logiciels sont en avance par rapport aux autres communautés de pratique.

3.1.4. Opérateurs

Les « opérateurs » appartiennent également au domaine de l'algorithmique. Dans les domaines de recherche liés à l'EIAH, ils apparaissent notamment dans la communauté Educational Data Mining (Romero et Ventura, 2007). Une forge logicielle est le lieu naturel adapté au cycle de vie complet de ces artefacts algorithmiques. Nous choisissons d'ajouter en conséquence une seconde forge logicielle. Les deux forges logicielles « processus d'analyse » et « opérateurs » peuvent être mise en commun dans l'architecture de la plateforme visée, cependant il semble important que les « opérateurs » puissent avoir une vie indépendante, hors toute activité de recherche expérimentale précise, une séparation des forges logicielles est donc aussi possible. Pour le choix pratique d'un cadre de

² « Une forge est un système de gestion de développement collaboratif de logiciel. » Wikipédia

développement des opérateurs, le critère principal sera la compatibilité avec celui fixé par l'environnement de définition des « processus d'analyse ».

3.1.5. Résultats

Enfin, les « résultats » terminent ce tour d'horizon des éléments architecturaux d'une plateforme de soutien à la recherche en EIAH. Les « résultats » apparaissent sous forme de données, parfois d'indicateurs ou de modèles. Ces résultats sont souvent mis en valeur pour donner lieu à diffusion au travers de visualisations, d'animations, de tableaux de bord, de publications. Pour la communauté Learning Analytics en particulier, l'aspect visuel et la lisibilité du résultat est importante, car elle doit être au service de l'humain (l'analyste, le décideur). Plus généralement, pour les « résultats », l'aspect communication est important. L'emploi de solutions logicielles de diffusion de l'information (CMS) est donc à ajouter. À nouveau l'ouverture et la facilité d'utilisation et d'extension de ces solutions logicielles sont un critère de choix pour déterminer quelle solution adopter.

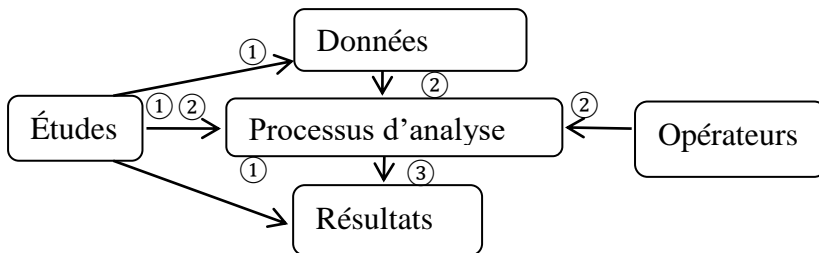


Figure 2. Architecture. Graphe des relations entre les artefacts. Un processus d'analyse est défini pour une étude ①, il utilise ② données, métadonnées et opérateurs pour produire ③ des résultats ; les données et résultats sont également définis pour une étude ①.

3.2. Architecture à 5 artefacts

Observons le graphe des relations entre ces cinq artefacts constituant le modèle de notre architecture de plateforme (cf. figure 2). L'étude sert de point d'entrée pour une partie des artefacts ; la notion de « processus d'analyse » est au centre de ce graphe. En effet, un processus d'analyse est défini pour une étude, il utilise des données et des opérateurs pour produire des résultats ; il est donc en relation avec l'ensemble des nœuds du graphe. Une étude est en lien direct avec les données, les analyses et les résultats, mais n'a pas de

relation directe avec les opérateurs utilisés lors de l'analyse. Les opérateurs sont indépendants de l'étude, ce qui justifie l'existence des deux cycles indépendants de (Mandran et al., 2015). Études et opérateurs sont les deux sources du graphe. Les résultats, quant à eux, apparaissent comme un produit final de l'activité. Pour la recherche, cependant, et le graphe ne le montre pas, chaque artefact peut également être compris comme un produit final : données, analyses, opérateurs ne sont pas seulement des moyens pour obtenir des résultats comme le montre ce graphe, ils sont aussi des objets de recherche et des résultats de la recherche. Ce graphe des relations est au cœur de l'architecture de la plateforme où chaque relation doit être représentée.

3.3. Ensemble de services associés à l'architecture

Passons à la description des services associés à cette architecture à cinq composants. L'analyse de l'activité de recherche expérimentale en EIAH à travers l'observation des cycles de vie des artefacts a abouti à une architecture à cinq éléments qui est significative de la complexité du projet. Pour l'élaboration de la liste des services disponibles pour la plateforme, l'objectif consistera à réduire cette complexité en tirant profit des similarités entre les cycles de vie observés et en bénéficiant du contexte d'un domaine d'application EIAH commun pour fournir une offre de services cohérents favorisant la mise en valeur de la transversalité des activités habituelles. Ainsi l'offre de service devra soutenir une vision orthogonale autant que possible indépendante des artefacts particuliers, pour renforcer les points communs de leur cycle de vie (conception, documentation, validation, etc.) et une méthodologie de travail de la recherche expérimentale plus générale (consultation, organisation, partage, etc.).

Un premier niveau de service global est en grande partie indépendant de l'activité de recherche en EIAH et concerne seulement la forme générale de la plateforme (web, partagé, etc.). Ce niveau général concerne l'ensemble de la plateforme et doit comporter un accueil général, une gestion des utilisateurs, de la documentation, des API pour fournir des arborescences de menu, des permaliens pour les pages, des statistiques de consultation, etc. L'un des points délicats à ce niveau est la gestion des utilisateurs et la définition des droits d'accès. La conservation des données, le domaine de la recherche, le respect de la vie privée, le travail collaboratif, etc. imposent chacun des contraintes qui globalement sont difficilement compatibles et posent des problèmes en soi non triviaux. Nous ne détaillerons pas ce niveau de services

généraux habituellement offert par les plateformes logicielles de développement de sites et applications web.

Aux niveaux suivants, la plateforme est divisée selon l'architecture en 5 sections, une par artefact. Chacune des 5 sections doit, autant que possible, être similaire aux autres (avec un découpage, décrit plus loin, similaire entre sections, avec une partie assurant une vision globale de l'ensemble des ressources disponibles dans la section, une partie associée aux éléments spécifiques du cycle de vie de l'artefact et une partie de description générale de chaque ressource avec les activités informelles et collaboratives associées). Pour chaque entité de l'architecture (étude, données, opérateurs, processus d'analyse et résultats) trois ensembles de services doivent ainsi être définis, en favorisant une séparation claire en une/des partie/s **indépendante/s** de l'artefact, et une partie comportant les services **spécifiques** de l'artefact. Nous proposons donc :

- un ensemble de services similaires pour tous les artefacts fournissant une **vision globale** des ressources d'un type donné (étude, donnée, opérateur, processus d'analyse ou résultat) disponibles sur la plateforme, avec l'élaboration de la liste des ressources présentes. Cette liste doit être consultable par l'utilisateur ou via une API pour une machine, interrogeable par recherche simple, en plein texte ou via les métadonnées. Cette liste doit être organisée selon des critères usuels (nom, date, auteur, etc.) et des critères liés au domaine des EIAH (discipline, niveau, type d'activité, etc.). Le résultat de la recherche doit mener à la présentation et la description complète de chaque ressource avec un permalien ou/et un identificateur unique. Cette partie doit être présentée de la même manière pour tous les artefacts, car essentiellement, c'est une liste de ressources, indépendante du type d'artefact considéré. Une illustration de cette partie est donnée plus loin, figure 3.

- un ensemble de **services spécifiques** prenant en charge la/les parties spécifiques du cycle de vie de chaque artefact, avec les phases de conception spécifiques (design, création, modification, amélioration, correction, etc.), les phases d'utilisation lorsqu'elles existent et qu'elles sont spécifiques (production, récolte, exécution, etc.) et les phases de validation dans la mesure où là aussi elles existent et sont spécifiques (contrôle, évaluation, etc.). Donnons quelques exemples de services spécifiques. Exemple de services spécifiques pour une « étude » : le moment de la rédaction d'un document de description d'une « étude » est important et spécifique, il doit expliciter des objectifs scientifiques à atteindre, des cadres de références épistémologique ou EIAH à prendre en compte, des

hypothèses expérimentales, les éléments éthiques à considérer, etc. un guide ou un service d'aide à la rédaction de ce document est souhaitable. Exemple de services spécifiques pour des « données » : l'import (ou l'export) de l'objet « données » demande une attention particulière, dans le cas de données massives c'est un point crucial. C'est un moment clé où, en plus d'obtenir l'objet, il faut s'assurer d'avoir aussi les caractéristiques de cet objet (format, métadonnées, etc.) et qu'un minimum de contrôle de qualité est nécessaire. Des services de téléchargements enrichis seront donc nécessaires. Exemple pour les « processus d'analyse » : le processus d'analyse est l'objet le plus complexe (au centre du diagramme figure 2) le plus singulier et le plus novateur de l'ensemble. C'est le moins évident à prendre en compte. Que ce soit son édition ou son exécution, les services à développer pour le prendre en charge représentent des défis majeurs pour le modèle de plateforme visée. Pour l'édition, l'objectif est de le séparer suffisamment du niveau algorithmique des opérateurs pour le rendre accessible aux utilisateurs. La programmation visuelle, et les services d'édition visuels associés, est la piste la plus prometteuse pour y arriver.

- un ensemble de services similaires pour tous les artefacts prenant en charge la **description générale** de l'artefact (métadonnées, contexte) et la part **informelle et collaborative de l'activité** de la recherche ainsi que quelques services spécifiques transversaux (gestion des questions éthiques et de l'anonymat ou des questions de la propriété intellectuelle et du partage, suivi des visiteurs). Dans cet ensemble, entrent en particulier la prise en compte des cycles de rétroaction et d'amélioration (forum de discussion, de suivi d'anomalies, évaluation par les pairs, gestion des versions, etc.), les diverses formes de production textuelle (publications, commentaires, documentation informelle, description narrative, etc.). À ce niveau peuvent aussi se mettre en place des services de mise en relation des différentes ressources entre elles. Ainsi, la référence à tel opérateur, dans un processus d'analyse, peut être exploité de manière automatique pour mettre en lien cet opérateur et ce processus dans la description de l'un et de l'autre. La mise à disposition de ces liens, leur publication, permettent, en complément des descriptions informelles, la navigation entre les différents artefacts en jeu à un moment donné et favorisent la sérendipité de la plateforme et des usagers. Cette partie doit être présentée de la même manière pour tous les artefacts, car essentiellement, il s'agit de services de gestion informelle indépendants du type d'artefact considéré. Une illustration de cette partie est donnée à la section suivante, figure 5.

Ensembles des services de la plateforme				
Services généraux de la plateforme (accueil, inscription, documentation, api, etc.)				
Services assurant une <u>vision globale</u> des études (affichage de l'ensemble des ressources, zone de recherche, etc.)	Services assurant une <u>vision globale</u> des données	Services assurant une <u>vision globale</u> des opérateurs	Services assurant une <u>vision globale</u> des analyses	Services assurant une <u>vision globale</u> des résultats
Services <u>spécifiques</u> du cycle de <u>vie</u> des études (guide préparatoires, zone de dépôt, etc.)	Services <u>spécifiques</u> du cycle de <u>vie</u> des données (visualisation, gestion des données, etc.)	Services <u>spécifiques</u> du cycle de <u>vie</u> des opérateurs (éditeur de code, forge, etc.)	Services <u>spécifiques</u> du cycle de <u>vie</u> des analyses (éditeur visuel, moteur d'exécution, forge, etc.)	Services <u>spécifiques</u> du cycle de <u>vie</u> des résultats (éditeur de texte, zone de dépôt, etc.)
Services de <u>description générale</u> d'une étude et des <u>activités informelles et collaboratives</u> associées à chaque étude (zone de partage, commentaires, etc.)	Services de <u>description générale</u> d'un ensemble de données et des <u>activités informelles et collaboratives</u> associées à chaque ensemble de données	Services de <u>description générale</u> d'un opérateur et des <u>activités informelles et collaboratives</u> associées à chaque opérateur	Services de <u>description générale</u> d'une analyse et des <u>activités informelles et collaboratives</u> associées à chaque de analyse	Services de <u>description générale</u> d'un résultat et des <u>activités informelles et collaboratives</u> associées à chaque résultat

Tableau 1 • Services. Outre les services communs généraux, chaque artefact (en colonne) est associé à des services similaires offrant une vision globale et une prise en compte de la part informelle et collaborative de l'activité, et à des services spécifiques de gestion du cycle de vie de l'artefact.

Ainsi, les cinq entités doivent être liées à trois ensemble de services (vision globale, cycle de vie, aspect informel et collaboratif) de manière transversale, selon un même schéma, réduisant l'apprentissage de l'ergonomie de la plateforme introduit par la complexité de l'architecture. Parmi l'ensemble de ces services à fournir, deux en particuliers sont à relever : 1) la prise en compte de la documentation informelle (à l'usage, c'est une évidence, nous manquons d'information pour décrire les travaux que nous menons, même à un niveau informel, les outils pourtant existent, il faut développer leur utilisation), 2) la mise en place des liens automatiques ou manuels entre les différents artefacts d'une recherche complexe selon les

relations données figure 1. : la séparation entre « études », « données », « opérateurs », « processus d'analyse » et « résultats » doit clarifier la présentation en séparant chaque élément, mais ne doit pas faire perdre l'ensemble, aussi des liens de navigation, type liens hypertextes doivent restituer la complémentarité de la recherche et permettre une description globale au travers des parcours de lecture, de liens en liens, qu'un observateur pourra en faire, favorisant ainsi au passage la sérendipité, voir plus loin figure 6.

Une idée de l'objectif visé peut être donnée par le prototype fonctionnel de plateforme décrit à la section suivante.

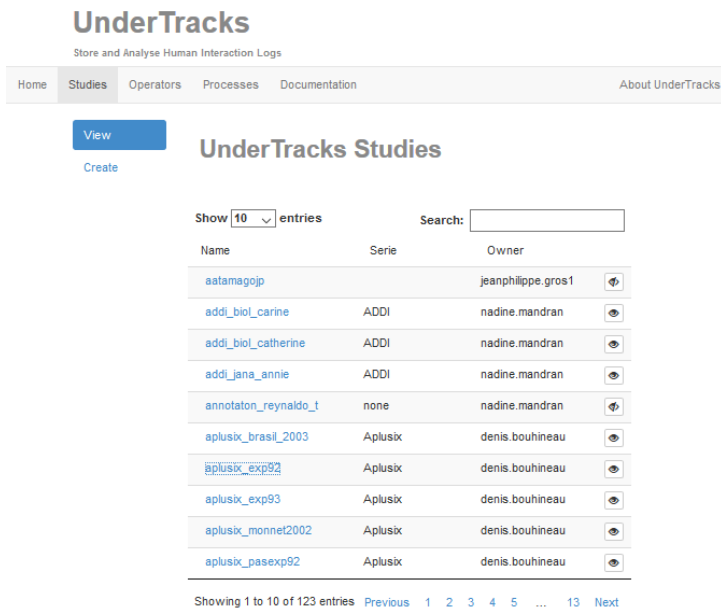


Figure 3. Vision globale des « études » sur la plateforme UnderTracks.

4. Premier prototype fonctionnel et résultats obtenus

Initié en 2010, le projet UnderTracks (Bouhineau *et al.*, 2013) avait pour objectif concret initial de constituer une plateforme de dépôt et de partage de traces EIAH seulement, avec la perspective de prolonger éventuellement, en cas de succès cette plateforme pour partager également des algorithmes calculant des indicateurs sur ces traces. En Juin 2016, une première version (V1.0) complète et fonctionnelle de cette plateforme a été obtenue et

déployée (<http://undertracks.imag.fr>) comprenant 1) une clarification sur la notion initiale de données, i.e. comportant une part « étude » et une part « données », 2) sur la notion initiale d'algorithme, i.e. comportant une part « opérateurs » et une part « processus d'analyse » et 3) laissant poindre la notion de « résultats ». La construction de cette première version complète et fonctionnelle d'une plateforme de soutien à la recherche en EIAH a permis la réflexion qui a donné lieu à la construction du modèle de cet article et peut servir d'outil support à l'exploration de ce modèle. UnderTracks V1.0 n'est pas un exemple parfait d'implémentation de la plateforme que nous proposons, l'ensemble du modèle n'est pas pris en compte, la séparation entre « étude » et « données » n'est pas encore assez claire, la notion de « résultats » n'est pas encore assez évidente. Le modèle de plateforme auquel nous sommes arrivé n'a été élaboré qu'au terme de la construction de UnderTracks V1.0, en constatant les usages possibles d'une telle plateforme. Ainsi, précurseur et premier prototype fonctionnel, UnderTracks permet d'observer ce que permettra une plateforme de soutien à la recherche expérimentale en EIAH.

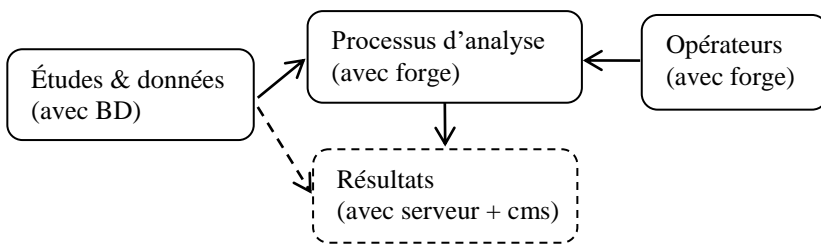


Figure 4. Architecture d'UnderTracks V1.0. Études & données sont fusionnées, processus d'analyse et opérateurs reposent sur une forge.

La mise en place effective des services sur UnderTracks V1.0 met en œuvre un découpage en trois parties, cf. figure 4., avec pour les « études », « opérateurs », « processus d'analyse » : une partie globale, une partie liée au cycle de vie de l'entité, une partie prenant en charge les aspects spécifiques. Cet ensemble de services est cohérent. L'expérience utilisateur est similaire pour les usages avec ces trois artefacts. Entre ces trois entités, la mise en place de liens automatiques de navigation, cf. figure 6, permet une navigation aisée. Pour autant l'appropriation de l'ensemble de la plateforme par chaque utilisateur n'est pas nécessairement l'objectif visé, l'organisation adoptée cherche plutôt à ce qu'un individu seul puisse travailler avec l'ensemble de la chaîne, sans devoir en maîtriser l'ensemble des chaînons, sinon celui qui le concerne. Une partie du défi sera réussie si chaque intervenant (l'expérimentateur produisant les données, l'informaticien construisant un

opérateur, l'analyste mettant en œuvre un processus d'analyse) arrive à s'approprier la partie de la plateforme qui le concerne et qu'un travail collaboratif entre les différents intervenants présents et utilisant ou adaptant les éléments mis à disposition par la communauté permet un travail de bout en bout.

The screenshot displays the UnderTracks web interface. At the top, the logo 'UnderTracks' is followed by the tagline 'Store and Analyse Human Interaction Logs'. A navigation bar includes links for Home, Studies, Operators, Processes, and Documentation, along with user information for 'About UnderTracks' and 'Hello Denis'. The main content area shows the breadcrumb 'Studies / View Existing / apusix_exp92' and a set of tabs for 'Events table', 'Users table', 'Context table', and 'Actions table'. The 'Study information' section lists details for 'apusix_exp92', including its beginning date (2002-2003), domain (algebra, didactic), learners level (3eme, 2nde, 1ere), Aplusix version (1), and studygroup (Aplusix). It also shows 'Total log entries: 443 240' and 'Reliability indicator: 1'. The 'Owner' is listed as 'denis.bouhineau@imag.fr'. Below this, a 'Viewers' section is empty, and 'Processes using it' is 'None for now'. The 'Study Description' section features the Aplusix logo and a reference to a paper by Bouhineau, D., Nicaud, J.F., Chaachoua, H., Bittar, M., and Bronner, A. (2005) titled 'Two years of use of the Aplusix system'.

Figure 5. Description générale d'une étude et des activités informelles et collaboratives associées à cette étude sur la plateforme UnderTracks,

Dans le détail des implémentations UnderTracks V1.0 comporte pour les infrastructures de haut niveau des développements communs en PHP, JavaScript, SQL avec la bibliothèque de présentation BootStrap et une base de données relationnelle postgresql. La partie « études » est portée par des développements spécifiques partageant les choix technologiques de la plateforme et comprend un dépôt de fichiers et une zone de commentaires libres (embryon de ce que pourraient donner une gestion électronique de documents et une zone d'échanges informels). La partie « données » repose essentiellement sur la base de données postgresql. La partie « opérateurs » repose essentiellement sur l'utilisation d'une forge logicielle publique FusionForge. La partie « processus d'analyse » est portée par les

développements de la partie « étude », l'utilisation d'une forge logicielle publique FusionForge et de la plateforme logicielle d'analyse Orange (Demsar *et al.*, 2013). L'examen des fichiers Orange des « processus d'analyse » permet une observation fine des données et opérateurs utilisés lors d'une analyse et permet la construction automatique des liens hypertextes « Etudes → Processus analyse », « Données → Processus analyse », « Opérateurs → Processus d'analyse ». Un gestionnaire de contenu, Drupal, est en place pour la communication de la plateforme et la publication des résultats obtenus. L'ensemble est visible à partir de l'adresse <http://undertracks.imag.fr>

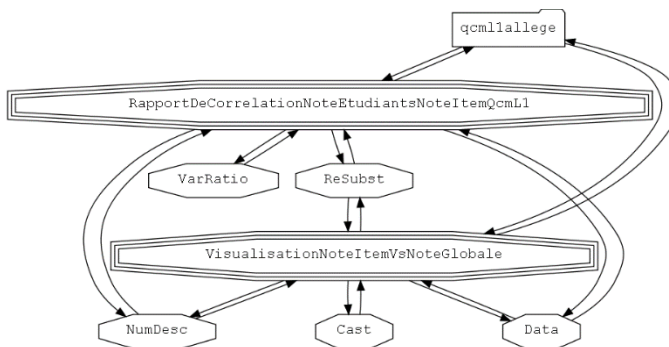


Figure 6. Extrait de la carte des liens du site UnderTracks V1.0. Visualisation des relations = lien hypertexte entre « études » (répertoire), « opérateurs » (octogones simples) et « processus d'analyse » (octogones triples).

Dès avant la version 1.0, la plateforme a pu accueillir, successivement au fil des développements : « études », « opérateurs » puis « processus d'analyse ». À ce jour, 121 « études », 52 « opérateurs » (auxquels s'ajoutent les opérateurs d'Orange) et 101 « processus d'analyse » sont disponibles. Parmi les « résultats », trois publications (Sanchez *et al.*, 2015), (Dessus *et al.*, 2016) et (Vermeulen *et al.*, 2016) ont été produites, en partie, à l'aide d'UnderTracks. Il est encore trop tôt pour faire une analyse des usages, mais des constantes apparaissent déjà. Les recherches en EIAH sont rarement ponctuelles et suivent des méthodologies communes liées à la discipline. Ainsi, pour les « études », les données déposées sur UnderTracks présentent souvent des caractéristiques récurrentes : présence de plusieurs études ou jeux de données relatives à une même recherche ; « études » réparties sur plusieurs années, parfois avec régularité (une étude par année). Ainsi se forment des familles d'études partageant des données ayant des tailles et des formes présentant une certaine similarité, parfois décrite en suivant la vie de l'EIAH et la vie de l'équipe de recherche (Chaachoua *et al.* 2007). Pour les

« processus d'analyses » les processus proches semblent moins définir des familles que des variantes (fork) d'un même processus, liées à des études différentes. Pour les « résultats », selon les analyses, les données utilisées étaient des données initiales ou des données secondaires, elles ont été enregistrées sur UnderTracks ; les opérateurs d'analyse de données et de visualisations présents sur la plateforme ont été mis à contribution ; des exports de données calculées ont été effectuée pour prolonger le travail sur d'autres plateformes de calcul ou de visualisation. Au niveau de l'activité de recherche, la combinaison des données et des analyses sur une plateforme commune ouverte a permis 1) une amélioration du travail de collaboration entre chercheurs ayant des métiers différents (didactique, science de l'éducation, informatique, statistique). 2) une amélioration de la qualité des données, des opérateurs et des analyses effectuées. Ainsi, selon l'artefact, le travail entre les acteurs a pu être collaboratif (par exemple pour la description des études et l'analyse des résultats) ou coopératif quand il y a eu une répartition des tâches selon les compétences et connaissances de chacun (par exemple pour les données vs les opérateurs et processus, entre les producteurs de donnée vs les programmeurs et analystes).

5. Conclusion et perspectives

Cette communication a présenté un modèle d'architecture et un ensemble cohérent de services associés, pour une plateforme de soutien à la recherche en EIAH pour la conception et l'échange de données, d'opérateurs, de processus d'analyse et de résultats issus d'études expérimentales. L'architecture est basée sur une analyse de l'ensemble de l'activité de recherche expérimentale en EIAH, elle met en avant et s'appuie sur un découpage fin de cette activité en artefacts décrits par leur cycle de vie qui étend (Romero et Ventura, 2007), (Stamper *et al.*, 2011), (Mandran *et al.*, 2015). L'ensemble de services associé permet de dépasser les modèles de plateformes d'archivage (Koedinger *et al.*, 2010), (Nelson et Warner, 2002), (Reffay *et al.*, 2012) pour prendre en compte la recherche en cours d'exécution, dans ces aspects non linéaires, informels et collaboratifs. Un prototype fonctionnel de ce modèle a été donné, fournissant une première preuve de concept de la possibilité de réaliser une plateforme de ce type.

L'avenir est à ce genre de plateforme permettant un renforcement de nos méthodologies de travail. Des guides de bonnes pratiques prenant en compte ces plateformes restent à prévoir pour tirer profit des divers partages envisagés, faciliter et enrichir les travaux en collaboration. Les premiers

retours ont montré que la réutilisation autorise des gains en efficacité, une plus grande fiabilité et réplicabilité des travaux. De nouveaux critères et indicateurs de qualités pour les données, opérateurs et analyses sont encore à penser pour suivre et porter ces évolutions. Avec l'apparition des premières plateformes EIAH fonctionnelles, comme UnderTracks V1.0, des analyses d'usages plus poussées sont également à organiser pour observer comment données et analyses s'organisent et si une particularité EIAH est visible dans les éléments ou les usages d'une telle plateforme. En particulier une des hypothèses initiales de la plateforme était la production d'opérateurs spécifiques aux EIAH, tenant compte de l'erreur de l'apprenant par exemple, mais pour l'instant cette hypothèse n'a pas été confirmée. En effet les opérateurs les plus utilisés sont des opérateurs d'analyse classiques.

Enfin, les difficultés pratiques pour la mise en place de telles architectures restent au niveau de l'étendue de l'activité et des artefacts à prendre en compte (le périmètre est vaste), l'intégration des environnements exogènes (de calculs, métiers, etc.) utilisés par ailleurs dans nos recherches, les multiples obstacles techniques, politiques et d'organisation avant d'arriver à un résultat effectivement utilisable (pour référence, la réalisation du prototype décrit en section 4 a pris plus de 5 ans et mobiliser autant (5) sinon plus de chercheurs).

De plus, la mise en place de la première version a montré la nécessité d'approfondir les questions éthiques (anonymat, propriété, type d'analyse acceptable vis-à-vis d'une communauté, etc.). Ces questions sont d'autant plus difficiles qu'elles sont relatives aux communautés d'usage et aux pays considérés et qu'elles évoluent dans le contexte actuel de la recherche.

Remerciements : Ce travail a été financé par l'ANR Hubble (ANR-14-CE24-0015).

Références

- (Anderson *et al.*, 2010)
Anderson J., Slater N., Lehnardt J.: CouchDB: The Definitive Guide. O'Reilly Media. (2010) 272p.
- (Bellanger et Tomassone, 2014)
Bellanger L. & Tomassone R. : Exploration de données et méthodes statistiques avec le logiciel R. Références sciences. Ellipses, 1st edition. ISBN 978-2-7298-8486-4. (2014)
- (Bouhineau *et al.*, 2013)
Bouhineau D., Luengo V., Mandran N., Ortega M. & Wajeman C. : Conception et mise en place d'un entrepôt de traces et processus de traitement EIAH : UnderTracks.

EIAH 2013 – 6e Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Toulouse, France. IRIT Press, (2013), 41-42.

(Chaachoua *et al.*, 2007)

Chaachoua, H., Croset, M.-C., Bouhineau, D., Bittar, M. & Nicaud, J.-F. : Description et exploitations des traces du logiciel d'algèbre Aplusix. Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF), 14, (2007).

(de Vries *et al.*, 2001)

de Vries, E., Pernin, J.-P. & Peyrin, J.-P. (eds.) : Hypermédias et Apprentissages 5. Actes du 5e colloque Hypermédias et Apprentissages. Coédition EPI-INRP (2001)

(Demsar *et al.*, 2013)

Demsar J., Curk T., Erjavec A., Gorup C., Hocevar T., Milutinovic M., Mozina M., Polajnar M., Toplak M., Staric A., Stajdohar M., Umek L., Zagar L., Zbontar J., Zitnik M. & Zupan B.: Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research 14. (2013), 2349–2353.

(Desmoulins *et al.*, 2001)

Desmoulins, C., Grandbastien, M. & Labat, J.-M (eds.) : EIAO2001 Environnements interactifs d'apprentissage avec ordinateur. Actes de la 6e conférence EIAO. Sciences et techniques éducatives, Vol. 8 — n° 1-2. (2001)

(Desmoulins *et al.*, 2003)

Desmoulins, C., Marquet, P. & Bouhineau, D. (eds.) : EIAH2003 Environnements Informatiques pour l'Apprentissage Humain. Actes de la 1ère conférence. INRP éd. (2003)

(Dessus *et al.*, 2016)

Dessus, P., Cosnefroy, O., & Luengo, V.: "Keep your eyes on 'em all!": A mobile eye-tracking analysis of teachers' sensitivity to students. In M. Sharples, K. Verbert, & T. Kloboučar (Eds.), Proc. 11th European Conf. on Technology Enhanced Learning, Lyon (2016).

(Frank *et al.*, 2016)

Frank E., Hall M.A. & Witten I.H.: The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, (2016).

(Koedinger *et al.*, 2010)

Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B. & Stamper, J.: A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press (2010)

(Mandran *et al.*, 2015)

Mandran N., Ortega M., Luengo V. & Bouhineau, D.: DOP8_Cycle: Merging both data and analysis operators life cycles for Technology Enhanced Learning. Conference LAK '15, the 5th International Learning Analytics and Knowledge Conference, Poughkeepsie, NY, (2015)

(Nelson et Warner, 2002)

Nelson, M. & Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting. Carl Lagoze C., Van de Sompel, H., (Eds.) Protocol Version 2.0. Document Version 2002/07/05. Open Archive Initiative. <http://www.openarchives.org/> (2002)

(Reffay *et al.*, 2012)

Reffay, C., Betbeder, M.-L. & Chanier, T.: Multimodal Learning and Teaching Corpora Exchange: Lessons learned in 5 years by the Mulce project. Special Issue on

dataTEL : Datasets and Data Supported Learning in Technology-Enhanced Learning, International Journal of Technology Enhanced Learning (IJTEL), Vol. 4 (1/2). (2012) 11-30

(Romero et Ventura, 2007)

Romero, C. & Ventura, S.: Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications Vol. 33 (1). (2007), 135–146.

(Sanchez *et al.*, 2015)

Sanchez E., Emin-Martinez V. & Mandran N. : Jeu-game, jeu-play, vers une modélisation du jeu. Une étude empirique à partir des traces numériques d'interaction du jeu Tamagocours. Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF), ATIEF, 22, (2015).

(Siegel et Retter, 2014)

Siegel E., Retter A.: eXist A NoSQL Document Database and Application Platform. O'Reilly Media. (2014) 562p.

(Stamper *et al.*, 2011)

Stamper, J.C., Koedinger, K.R., Baker, R.S.J.d, Skogsholm, A., Leber, B., Demi, S., Yu, S. & Spencer, D.: Managing the Educational Dataset Lifecycle with DataShop. Artificial Intelligence in Education. G. Biswas, S. Bull, J. Kay, and A. Mitrovic, (Eds). Springer Berlin Heidelberg. (2011) 557–559.

(Vermeulen *et al.*, 2016)

Vermeulen M., Mandran N. & Labat J.M.: Chronicle of a scenario graph: from expected to observed learning path. In M. Sharples, K. Verbert, & T. Klobučar (Eds.), Proc. 11th European Conf. on Technology Enhanced Learning, Lyon (2016).