

# HUBBLE, UN OBSERVATOIRE DES ANALYSES DES TRACES D'APPRENTISSAGE ANR 14 CE24 0015 (2015-2018)

Responsable Scientifique Vanda Luengo, LIP6, SU

Laboratoires : Laboratoires : LIP 6 Vanda Luengo, LIRIS Nathalie Guin, LIG Denis Bouhineau, IFE Philippe Daubias, STEF Eric Bruillard, LIUM Sébastien Iksal, labSTICC Serge Garlatti, LINA Yannick Prié

Entreprise : OPENCLASSROOM Romain Kuzniak

## A.1 RESUME

L'objectif d'HUBBLE est de créer un observatoire national pour la construction et le partage de processus d'analyse de données massives, issues des traces des environnements de type e-learning. L'enjeu principal est de permettre la conception, la capitalisation et la réutilisation de ces analyses en associant des acteurs de différents domaines. HUBBLE a vocation à fédérer des équipes de recherche dans plusieurs domaines de l'informatique et des sciences humaines. Le projet considère les problèmes éthiques en proposant un comité éthique et des outils.

HUBBLE est organisé autour de deux concepts : (1) un *cas d'étude* est une situation d'apprentissage particulière, avec des enseignants, des apprenants, des gestionnaires et une plateforme d'apprentissage, (2) un *scénario d'analyse* décrit les objectifs de l'analyse, le contexte d'usage et les indicateurs ou modèles attendus.

Les productions scientifiques du projet sont de plusieurs ordres : D'abord, des outils et méthodes permettant, pour les analystes non spécialisés en analyse de données et non informaticiens, d'explorer rapidement des analyses et des traces ; De plus, des méthodes de conception dynamique afin de proposer des tableaux de bord adaptés aux besoins des enseignants. Ensuite pour la communauté des Learning Analytics (LA), chercheurs et analystes, HUBBLE offre la possibilité, à travers des formalismes et outils, de capitaliser les analyses, en les décrivant de manière fine (justification des choix) et indépendamment d'une plateforme d'analyse, pour ensuite les consulter et les réutiliser. La dimension éthique a été pleinement intégrée dans la conception des analyses en tenant compte non seulement de l'éthique sur les données mais aussi de l'éthique sur les analyses.

## A.2 ENJEUX ET PROBLEMATIQUE, ETAT DE L'ART

Dans le domaine du e-learning, la collecte et l'analyse de traces sont des domaines émergents représentés par la communauté d'Educational Data Mining (EDM - <http://educationaldatamining.org/>) et de Learning Analytics (LAK - <https://solaresearch.org/events/lak/>). Plusieurs travaux s'intéressent à la capitalisation de traces. DATAShop [Koedinger et al 2010] est l'une des plateformes les plus connues et utilisée en EDM. Elle met à disposition des traces avec l'objectif spécifique de trouver et tester des modèles d'apprenant. Il s'agit de traces d'interaction avec des tuteurs intelligents de type cognitive tutor.

En France, le projet MULCE [Reffay et al. 2012] est représentatif de l'effort autour de la collecte et du partage de traces. Ce projet propose la diffusion de corpus d'apprentissage, appelés LETEC, avec comme objectif la mise à disposition non seulement des données sur les résultats d'une formation, mais également des données sur le contexte, c'est-à-dire caractérisant le dispositif de formation ainsi que le dispositif de recherche. L'objectif est de rendre ce type de données visibles, partageables et réutilisables.

Ces deux projets ne capitalisent pas les processus d'analyse mais proposent des liens vers des outils libres d'analyse développés par des chercheurs. Les auteurs de ces projets soulignent l'importance de réaliser des connections entre leur plateforme et ce type d'outils. Ainsi, il existe des plateformes de traitement de traces génériques (comme Weka pour le data mining ou R pour la statistique), mais certains auteurs [Romero et al. 2007] pointent la nécessité de concevoir des processus d'analyse propres au domaine d'e-learning.

De ce point de vue (outils d'analyse), la différence principale des deux communautés (EDM et LAK) sont les cibles des productions. En effet, les algorithmes et outils produits dans EDM sont principalement utilisés par les systèmes informatiques alors que la communauté LAK s'intéresse principalement à fournir des outils pour les utilisateurs humains du domaine. De plus, EDM met l'accent sur la découverte automatique des connaissances alors que LAK met l'accent sur la découverte à partir du jugement humain [Siemens et Baker 2012].

Dans le cadre du projet HUBBLE, nous avons souhaité associer ces deux approches. Le consortium du projet associe ainsi des équipes qui sont parmi les acteurs principaux dans le domaine des EIAH et qui s'intéressent à l'usage des plateformes e-learning ainsi qu'à la collecte et l'analyse des traces d'e-learning.

De plus, au démarrage du projet les expérimentations autour des Learning Analytics en France étaient peu nombreuses et les chercheurs concernés, relativement peu habitués à prendre en considération les questions

éthiques liées aux traitements typiques des Learning Analytics. Le projet avait donc pour mission de définir un cadre éthique et juridique pour encadrer les risques liés à l'utilisation des Learning Analytics.

### **A.3 APPROCHE SCIENTIFIQUE ET TECHNIQUE**

HUBBLE est organisé de façon itérative et bottom-up autour de deux concepts : le cas d'étude et le scénario d'analyse. Il y a deux types de scénario d'analyse : les scénarios chercheurs permettant de répondre à une question de recherche et les scénarios décideurs (enseignant, apprenant, responsable formation, ...) permettant l'accompagnement à la prise de décision liée à l'apprentissage [Luengo 2014]. Les scénarios d'analyse sont améliorés à partir des évaluations des cas d'études pour lesquels ils ont été conçus et à partir des nouveaux cas d'études pour lesquels ils sont réutilisés. Grâce au consortium, nous avons pu être opérationnels depuis le démarrage du projet avec plusieurs cas d'études fournis par ses membres. Le projet a mis en place trois itérations.

Les lots 1 et 2 du projet ont visé la description des cas d'études et des analyses. Après une première spécification faite séparément, avec une grille de description pour les cas d'étude et une autre pour les scénarios d'analyse, nous avons constaté qu'il était nécessaire de tenir compte des deux dimensions ensemble. Nous avons décidé donc de fusionner les deux lots. Il a fallu également proposer un lexique qui établisse un vocabulaire commun [Mandran et al. 2015, Sanchez et al. 2017].

Le lot 3 visait à modéliser, importer et collecter en temps réel les traces. Il a fallu choisir et instancier un format de traces adapté à cet objectif.

Le lot 4 visait à mettre en œuvre les scénarios d'analyse. Dans un premier temps, des analyses ont été effectuées par des sous-groupes de partenaires autour de quelques cas d'étude. Chaque sous-groupe développait un ou plusieurs scénarios dans une liste de quatre, en découvrant et utilisant les plateformes des partenaires. Cette première itération a permis de voir des limites dans l'utilisation des plateformes et dans la capitalisation des processus d'analyse et qui ont ensuite orienté le travail du lot 4.

Le lot 5 a adopté une démarche itérative et expérimentale pour la définition d'un processus complet de conception de tableaux de bord d'apprentissage. La première itération a été réalisée à partir de l'ensemble des cas d'utilisation de HUBBLE pour ensuite se concentrer autour de quelques cas. Elle a également permis de définir un premier prototype de génération de tableaux de bord d'apprentissage [Dabbebi et al 2017]. La deuxième itération a permis de mieux identifier les dimensions visuelles des tableaux de bord, et d'organiser les éléments contextuels principaux. La troisième itération a permis de valider une approche participative [Dabbebi et al 2018] de capture des besoins liés aux cas d'utilisation et la génération dynamique de tableaux de bord d'apprentissage. Le lot 6 a constitué dès fin 2016 un "Comité d'Éthique pour les Learning Analytics" (CELA) qui a conduit dès 2017 à la proposition d'un formulaire-guide de "Soumission d'une recherche au CELA". Ce formulaire a été rempli par les responsables des données de chaque cas dans HUBBLE et validé par les responsables du lot 6. Le protocole prévoyait deux niveaux de validation possibles : le premier niveau par les responsables du lot éthique, le deuxième en réunissant le CELA si nécessaire. Dans le cadre du projet HUBBLE, la mobilisation du premier niveau a suffi pour valider l'approche proposée par les chercheurs responsables des données.

### **A.4 RESULTATS OBTENUS**

Les avancées scientifiques dues au projet HUBBLE sont de rendre les analyses plus accessibles aux différents acteurs :

- pour les analystes non spécialisés en analyse de données et non informaticiens, qui peuvent maintenant explorer rapidement les cas d'étude disponibles (livrables du lot 1 et 2) et qui peuvent modéliser et visualiser des traces, les explorer, effectuer des transformations pour calculer des indicateurs et construire des connaissances (livrables des lots 3 et 4)
- pour la communauté des LA (chercheurs et analystes), HUBBLE offre la possibilité de capitaliser les analyses, en les décrivant de manière fine (justification des choix) et indépendamment d'une plateforme d'analyse, pour ensuite les consulter et les réutiliser (livrables du lot 4).
- pour les enseignants via des tableaux de bord adaptés à leurs besoins (livrables du lot 5).

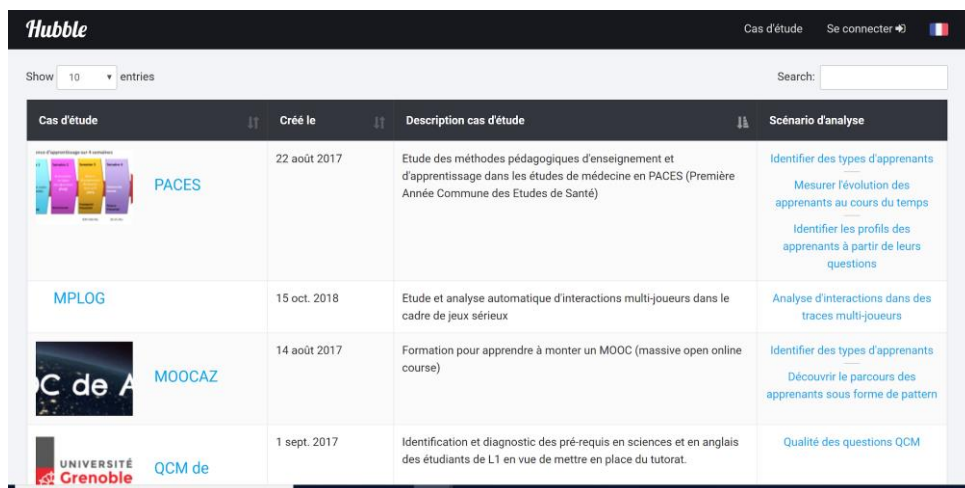
De plus, sur les usages des LA en général, la dimension éthique a été pleinement intégrée dans la conception des analyses (livrables du lot 6).

Enfin, ces usages passent également par la diffusion et la médiation scientifique dans laquelle l'équipe HUBBLE s'est fortement impliquée à travers des actions avec la Direction du Numérique pour l'Éducation (DNE), les rapports de vulgarisation scientifique, le travail de terrain auprès des utilisateurs finaux et la présentation des concepts et méthodes dans des conférences et séminaires de vulgarisation et médiation scientifique.

## GESTION DE CAS D'ETUDES ET SPECIFICATION DES SCENARIO D'ANALYSE

Les réflexions des lots 1 et 2 centrés sur les cas d'étude et scénarios d'analyse ont abouti à un catalogage de scénarios d'analyse et la production de synthèses méthodologiques.

Après une ébauche de catalogue indexé sur des analyse-type sous la forme d'un site Github, qui s'est révélée difficile d'usage pour certains types d'utilisateurs finaux, un catalogue centré sur des cas d'études a été finalement retenu pour le projet Hubble (Figure 1).



The screenshot shows the Hubble Catalogue interface. At the top, there is a search bar and a 'Se connecter' button. Below the search bar, there is a table with four columns: 'Cas d'étude', 'Créé le', 'Description cas d'étude', and 'Scénario d'analyse'. The table contains four rows of data, each representing a different study case and its associated analysis scenario.





Cas d'étude	Créé le	Description cas d'étude	Scénario d'analyse
 PACES	22 août 2017	Etude des méthodes pédagogiques d'enseignement et d'apprentissage dans les études de médecine en PACES (Première Année Commune des Etudes de Santé)	Identifier des types d'apprenants Mesurer l'évolution des apprenants au cours du temps Identifier les profils des apprenants à partir de leurs questions
 MPLOG	15 oct. 2018	Etude et analyse automatique d'interactions multi-joueurs dans le cadre de jeux sérieux	Analyse d'interactions dans des traces multi-joueurs
 MOOCAZ	14 août 2017	Formation pour apprendre à monter un MOOC (massive open online course)	Identifier des types d'apprenants Découvrir le parcours des apprenants sous forme de pattern
 QCM de	1 sept. 2017	Identification et diagnostic des pré-requis en sciences et en anglais des étudiants de L1 en vue de mettre en place du tutorat.	Qualité des questions QCM

Figure 1. Interface du Catalogue Hubble

Des documents synthétiques sous forme de revues bibliographiques ont été produits [Sanchez et al. 2017], avec pour fonction d'homogénéiser le vocabulaire utilisé entre sciences humaines et informaticiens, en reprenant les taxonomies issues de la littérature pour décrire les techniques d'évaluation d'EIAH, notamment pour les tableaux de bords, les méthodes mixtes, ou les traitements statistiques des données.

Enfin, les premières étapes du projet HUBBLE ont vu l'apparition d'un vocabulaire ad hoc pour décrire des méthodes de traitement de données identifiées dans la littérature sous des noms variés (association rule mining, etc.), et nous avons réintroduit cette terminologie dans les échanges entre partenaires, par un document recensant la terminologie généralement employée. Une démarche similaire a été menée avec les méthodes mixtes [Mandran et al 2015], visant à faire le pont entre sciences humaines et informatique. Une typologie de méthodes mixtes a été produite à partir de la littérature et illustrée avec des analyses fondées sur des traces d'interaction.

## METHODES ET MODELES DE COLLECTE DES TRACES

En lieu et place d'un format commun pour les traces initialement prévu, nous avons choisi pour le lot 3 de nous appuyer sur un format de traces reconnu et dédié aux EIAH, TinCan/xAPI, pour ensuite l'enrichir par un vocabulaire émergent des traces du projet HUBBLE. La spécification xAPI propose, en plus d'un vocabulaire, un système de gestion de traces, nommé *Learning Record Store* (LRS). Nous avons donc instancié un LRS commun aux différentes équipes pour y déposer les traces étudiées au sein de HUBBLE. Ce LRS permet d'échanger les traces entre les différentes plateformes d'analyse existant au sein du projet (KTBS, Undertracks, UTL...), pour lesquelles des passerelles avec xAPI ont été développées. Afin de téléverser les traces existantes vers le LRS, un outil de séquençage a été développé. Cet outil permet de faire correspondre les champs de la trace sous forme tabulaire au vocabulaire des *statements* xAPI. Chaque ligne du fichier de trace CSV correspond à un unique *statement*. Un minimum de connaissance dans le format xAPI est requis pour utiliser cet outil. Parallèlement au LRS, deux documents permettent de référencer d'une part la trace téléversée, à savoir son identifiant, les conditions de téléversement, etc. ; et d'autre part le vocabulaire qu'il a fallu ajouter au vocabulaire xAPI. Un autre avantage de xAPI est qu'il supporte nativement la collecte dynamique (c'est à dire pendant l'activité tracée plutôt qu'*a posteriori*), ce qui était un objectif du lot 3.

## METHODES ET OUTILS POUR LA MISE EN ŒUVRE DE SCENARIOS D'ANALYSE

Trois axes ont été principalement exploités dans cette partie, la mise en œuvre de plusieurs scénarios d'analyses, des développements sur les plateformes Orange-UT et kTBS4LA et une thèse en informatique sur la problématique de la capitalisation des processus d'analyse de traces d'apprentissage au sein de la communauté des Learning Analytics (LA).

Pendant le projet, les partenaires ont développé, dans les plateformes d'analyse existantes au sein des laboratoires, des outils à destination des analystes non informaticiens. Deux plateformes étaient concernées : Orange-Undertracks [Bouhineau et al. 2018], développée au LIG, permettant aux utilisateurs de définir des processus d'analyse graphiquement, sous la forme de diagrammes de flux de données et kTBS4LA, un outil d'analyse construit sur la plateforme kTBS développée au LIRIS [Casado et al. 2017]. Ce dernier permet à un utilisateur de définir un modèle des traces qu'il souhaite analyser, de visualiser ces traces pour les explorer, d'effectuer des transformations de traces et de calculer des indicateurs sur l'activité des apprenants. Les spécificités de cet outil sont l'importance donnée à la temporalité des traces, une représentation explicite de la sémantique des interactions tracées, et la possibilité pour l'utilisateur de manipuler directement les données qu'il analyse. Une collaboration avec les analystes de l'IFé a permis de faire évoluer l'outil dans une démarche de conception participative, afin qu'il satisfasse au mieux les besoins des analystes (Bonvin et al. 2018).

Pour la thèse d'Alexis Lebis, il s'agissait de permettre de partager, adapter et réutiliser des processus d'analyse de traces. Actuellement, cette capitalisation est limitée par deux facteurs importants : les processus d'analyse sont dépendants des outils d'analyse qui les mettent en œuvre - leur contexte technique - et du contexte pédagogique pour lequel ils sont menés. Cela empêche de les partager, mais aussi de les ré-exploiter simplement en dehors de leurs contextes initiaux, quand bien même les nouveaux contextes seraient similaires.

La thèse a donné lieu à quatre contributions :

- La formalisation d'un cycle d'élaboration et d'exploitation des processus d'analyse, afin d'en définir les différentes étapes, les différents acteurs et leurs différents rôles. Cette formalisation est accompagnée d'une définition de la capitalisation et de ses propriétés.
- Un méta-modèle qui permet de décrire les processus d'analyse indépendamment des outils d'analyse. Il formalise la description des opérations utilisées dans les processus d'analyse, des processus eux-mêmes et des traces utilisées, afin de s'affranchir des contraintes techniques occasionnées par ces outils. Ce formalisme commun aux processus d'analyse permet aussi d'envisager leur partage. Il a été mis en œuvre et évalué dans un prototype [Lebis et al 2016].
- Un framework ontologique pour les processus d'analyse, qui permet d'introduire de manière structurée des éléments sémantiques dans la description des processus d'analyse. Cette approche narrative enrichit ainsi le formalisme précédent et permet de satisfaire les propriétés de compréhension, d'adaptation et de réutilisation nécessaires à la capitalisation. Cette approche ontologique a été mise en œuvre et évaluée dans un prototype [Lebis et al 2017, 2018].
- Des mécanismes d'assistance aux acteurs, notamment une nouvelle méthode de recherche des processus d'analyse, s'appuyant sur les propositions précédentes. Le cadre ontologique de l'approche narrative est exploité pour définir des règles d'inférence et des heuristiques permettant de raisonner sur les processus d'analyse dans leur ensemble (e.g. étapes, configurations) lors de la recherche. Le réseau sémantique sous-jacent à cette modélisation ontologique est également utilisé pour renforcer l'assistance aux acteurs en leur fournissant des outils d'inspection et de compréhension lors de la recherche. Cette assistance a été mise en œuvre dans un prototype, et évaluée empiriquement [Lebis et al 2018].

De plus, nous avons produit des scénarios d'analyse qui sont disponibles dans le catalogue HUBBLE. Certains sont des scénarios décideurs et certains des scénarios chercheurs. Tous les scénarios chercheurs ont donné lieu à des publications scientifiques (Cisel 2018, Minho et al. 2016, Bonvin et al. 2018, Sanchez et al 2015, Sanchez et al. 2017, Pierrot et al. 2017, El-Kechaï et al 2016, Harrak et al 2018, Harrak et al. 2019, Sanchez 2017, Pigeau et al 2019). Les scénarios décideurs ont conduit à quelques publications (Bey et al. 2018, Hosseuin et al. 2019, Itani et al. 2018, Luengo et al. 2019) et ont été livrés aux utilisateurs. Cependant, ils ne sont pas nécessairement introduits dans les usages de ces utilisateurs finaux.

Pour illustrer ce dernier point, on peut citer la collaboration du laboratoire LS2N avec l'entreprise OpenClassrooms, disponible dans le catalogue. La première partie de ce travail a consisté à étudier les données OpenClassrooms, avec pour résultat (1) un état de l'art sur les indicateurs utilisés dans le domaine des cours en ligne, pour de la prédiction ou du profilage d'apprenants ; (2) une analyse visuelle des indicateurs sur différents profils d'apprenants (abandon, premium/non premium, succès/échec, ...) ; (3) une analyse séquentielle des données, basée sur des approches de pattern mining et de process mining, l'objectif étant de rechercher des

différences sur la manière dont les apprenants parcourent un cours ; (4) un état de l'art sur les indicateurs utilisés dans le domaine des cours en ligne, pour de la prédiction ou du profilage d'apprenants ; Une deuxième partie a permis de déterminer la meilleure méthode de prédiction d'échec des apprenants, à savoir par régression logistique et AdaBoost. Ces résultats ont été communiqués aux décideurs.

## CONCEPTION DES TABLEAUX DE BORD

Les travaux menés dans le cadre du lot 5 avaient pour but de proposer des modèles opérationnalisés permettant la génération dynamique de tableaux de bord d'apprentissage. Pour cela, les différents cas d'études ont été analysés et une démarche participative a été entreprise. Dans un premier temps, un questionnaire a été élaboré (L5.1b - Grille d'expression des besoins pour les tableaux de bord) et a permis d'obtenir un certain nombre de réponses (Dabebbi, 2017).

Suite à cela, le constat a été fait que l'expression de besoins d'observation par les usagers était très compliquée en raison des difficultés de perception de ce que serait un tableau de bord. Le cas d'étude TactileoMap, nous a permis de travailler directement avec les usagers sur des sessions de MockUp et par la suite d'identifier que le nœud central était la décision à prendre à l'aide du tableau de bord. La démarche participative a donc évolué et un outil de capture des besoins sous la forme d'un espace de conception avec des cartes de décision, de données et de visualisation a été proposé (Gilliot et al., 2018, <http://perso.univ-lemans.fr/~siksai/HUBBLE/L5.5-KitConceptionTB.zip>). Ce dernier a été testé dans le cadre de HUBBLE mais aussi dans divers séminaires de l'Education Nationale (comme les rencontres de l'Orme) ainsi que dans des projets de recherche obtenus suite au projet HUBBLE (Les projets incubateurs de la DNE), directement avec des équipes pédagogiques au sein d'établissements scolaires.

Un processus de génération dynamique de tableaux de bord d'apprentissage contextuels a été proposé (Dabebbi et al., 2017) et outillé avec différents modèles (L5.3 et L5.4 – Modèles et Génération de Tableaux de bord) utilisés par un prototype. Les modèles et le prototype ont évolué au fil des rencontres avec les usagers de TactileoMap, jusqu'à la validation des tableaux de bord générés au profit de cette équipe (Dabebbi et al., 2019).

## ÉTHIQUE DANS LES LA

À travers HUBBLE, les chercheurs du lot 6 ont pu sensibiliser l'ensemble des chercheurs impliqués dans les différents cas d'étude de HUBBLE et en premier lieu les responsables des données qui ont dû répondre aux questions du formulaire guide et ainsi prendre conscience ou même préciser les risques liés au traitements des données, l'usage ou le détournement qui pourrait en être fait, et par conséquent préciser les règles de sécurisation et d'anonymisation des données collectées, et des résultats fournis par les traitements. De plus des outils d'accompagnement ont été produits (May et al. 2016, May et al. 2017, Reffay et al. 2017, Mille et al. 2017).

## A.5 EXPLOITATION DES RESULTATS

Les avancées et résultats du projet HUBBLE ont donné lieu à plusieurs types d'exploitation. D'abord, les résultats ont permis la mise en place de nouvelles collaborations scientifiques au niveau national (labex Cominlabs Rennes, Rectorat de Nantes avec le TECHNE et le LIUM, Rectorat d'Orléans avec le LIUM, le LABSTICC et TECHNE), européen (projet Erasmus avec Le LIP6 et Nancy) et internationale (le LIRIS et Georgia Tech, le LIP6 avec l'université University Hall, Singapour).

Ensuite, HUBBLE a clairement participé à l'introduction de la thématique de recherche des Learning Analytics (LA) en France. La synergie avec le réseau Orphée a permis une animation scientifique autour des questions relatives aux LA. Le projet devient une référence nationale comme en témoignent les différentes manifestations auxquelles nous étions invités à participer au nom du projet. La thématique de la conférence francophone du domaine, EIAH 2019, est sur les LA.

De plus, les efforts de médiation auprès de l'éducation nationale, à travers la Direction du Numérique pour l'Éducation, permettent la sensibilisation des usages des LA, et des projets de terrain sont actuellement mis en place.

Enfin, des formations dans lesquelles sont impliquées les équipes du projet HUBBLE ont mis en place des cours sur les LA au niveau des masters, mais aussi pour les ingénieurs pédagogiques dans les universités. Ces formations utilisent notamment le catalogue HUBBLE comme guide méthodologique.

Les formulaires renseignés par chaque cas de HUBBLE constituent une trace concrète du sérieux avec lequel la dimension éthique a été prise en considération par les chercheurs du projet. La question de l'anonymisation des traces laissées dans les plateformes de téléformation n'a été traitée que pour les données de type traces d'actions enregistrées par le système. La démarche, le formulaire et ses résultats ont été rendus accessibles via

le site du projet HUBBLE et ont été partagés avec la communauté scientifique au-delà de l'hexagone et au-delà du projet. Le questionnaire auquel invite le formulaire a par exemple été réutilisé dans la mise en œuvre d'un jeu sérieux en anatomopathologie (indépendant de HUBBLE) au sein de l'université de Franche-Comté pour encadrer les traces et leurs usages par les différentes parties prenantes (étudiants, enseignants, concepteurs, administrateurs). Les résultats liés à l'éthique sont également exploités dans d'autres projets. À titre d'exemple, le projet ANR Pastel et le projet PIA escri+ nécessitent la mise en place d'une démarche similaire à celle du projet HUBBLE. La charte éthique du projet HUBBLE est également utilisée dans le cadre du projet ANR COMPER.

## A.6 DISCUSSION

HUBBLE est un projet qui associe de nombreux partenaires et qui est pluridisciplinaire. La démarche par itération a montré la nécessité de s'organiser de façon à garantir l'implication de tous les rôles nécessaires selon le scénario d'analyse (experts du cas d'étude, analyste, décideur, utilisateur ciblé et/ou chercheur).

De plus, les scénarios d'analyse ne décrivent pas de façon univoque leurs objectifs et il est nécessaire de les associer au contexte. Ainsi, par exemple, la qualité d'un QCM ne se mesure pas de la même façon pour un cas d'étude de première année de médecine (cas d'étude PACES) que pour un cas d'étude de première année de mathématique et informatique (Cas d'étude Grenoble pré requis UGA ou LIFAsker).

Les travaux éthiques du projet HUBBLE ont eu lieu conjointement à la mise en œuvre de la RGPD au niveau européen et ont permis aux chercheurs de HUBBLE de mieux s'y préparer. Il reste cependant beaucoup à faire en particulier sur les traitements semi-automatiques d'anonymisation. En effet, pour toutes les productions (devoirs, documents multi-média, infographies, etc.) et interactions orales (audio ou visio-conférence) ou écrites (chat, forum, messageries), le projet HUBBLE n'a pas pu développer plus avant les techniques d'anonymisation. Il a été montré par de nombreux exemples que ces techniques ne pouvaient être entièrement automatisées.

De façon plus générale, les LA commencent à s'installer dans le paysage scientifique mais également dans sa possibilité d'usage dans le monde éducatif. Les véritables usages et comment cela peut accompagner la prise de décision pour améliorer les apprentissages restent des questions ouvertes.

## A.7 CONCLUSIONS

Au terme du projet, il ressort que l'analyse des données requiert la combinaison de compétences diverses et complexes permettant :

- une compréhension poussée des environnements informatiques utilisés (LMS, jeux, simulateurs) et des choix effectués pour leur mise en place, par exemple les règles du jeu mis en place ou les contraintes de conception des activités et ressources dans un LMS.
- une connaissance fine du contexte dans lequel les environnements informatiques sont utilisés. Il s'agit de connaître les objectifs poursuivis par l'enseignant, les contraintes institutionnelles et hiérarchiques auxquelles l'enseignant est soumis et les élèves impliqués.
- une connaissance des données collectées, de leur format et de leur structure. Cela a requis pour certains cas des itérations de conception pour produire des nouvelles traces.
- une connaissance des outils de traitement et de visualisation et la capacité à les développer ou les adapter aux besoins de l'analyste ;
- une maîtrise des modèles théoriques sur lesquels se fondent les analyses.

De plus, au regard des types de scénarios identifiés dans le projet (scénarios chercheurs et scénarios décideurs), les méthodes et itérations ne sont pas du même ordre. Il apparaît qu'une démarche centrée tableaux de bord est plus appropriée pour les scénarios décideurs alors qu'une démarche centrée contexte et données est plus appropriée pour les scénarios chercheurs.

Enfin, le projet n'a pas permis, par la composition de l'équipe et la genèse du projet, de pouvoir aller plus loin dans l'appropriation des LA dans un contexte véritable d'enseignement pour pouvoir analyser les bénéfices d'un tel usage et ainsi pouvoir fermer la boucle des LA.

## A.8 REFERENCES

- Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R., Aleven, V., Cocea, M., Hershkovitz, A., de Carvalho, A.M.J.B., Mitrovic, A., Mathews, M. (2013) Modeling and Studying Gaming the System with



Educational Data Mining. In Azevedo, R., & Aleven, V. (Eds.) *International Handbook of Metacognition and Learning Technologies*. pp. 97-116. New York, NY: Springer.

- Bey, A., Bouhineau, D. (2018) "Human Scoring versus Automatic Scoring of Computer Programs: Does Algo+ score as well as instructors ? An Experimental Study." The IEEE Computer Society Conference Publishing Services. 18th IEEE International Conference on Advanced Learning Technologies, ICALT2018, Mumbai, India. 18th IEEE International Conference on Advanced Learning Technologies Conference Proceedings. <hal-01843982>
- Bonvin, G., Sanchez, E., Champin, P., Casado, R., Guin, N., Lefevre, M. (in press). "Assessing Social Engagement in a Digital Role-Playing Game: Changes over Time and Gender Differences." In R. Bottino & D. Passey (Eds.), *Empowering Learners for Life in the Digital Year*, proceedings of the OCCE 2018: Springer.
- Bouhineau, D., Luengo, V., Mandran, N. (2018) "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." <hal-01843970>
- Casado, R., Guin, N., Champin, P-A., Lefevre, M. (2017) "ktBS4LA : une plateforme d'analyse de traces fondée sur une modélisation sémantique des traces." In ORPHEE-RDV 2017, Atelier "Méthodologies et outils pour le recueil, l'analyse et la visualisation de traces numériques d'interaction."
- Cisel, M., Mano, M., Bachelet, R., Silberzahn, P. (2015) "A Tale of Two MOOCs: Analyzing Long-Term Course Dynamics." European Moocs Stakeholders Summit (eMOOCs), Mons, Belgium.
- Dabbebi, I. (2016) "Modélisation des tableaux de bord contextuels, adaptatifs et dynamiques." RJC-EIAH2016.
- Dabbebi, I., Iksal, S., Gilliot, J.-M., May, M., Garlatti, S. (2017) "Towards Adaptive Dashboards for Learning Analytic: An Approach for Conceptual Design and implementation." CSEDU 2017.
- Dabbebi, I., Gilliot, J. M., Iksal, S. (2019). "User Centered Learning Analytics Dashboard Generation." In 11th International Conference on Computer Supported Education (CSEDU 2019).
- Fuchs, B. (2017) "Extraction d'épisodes séquentiels à partir de Traces : application au jeu Tamagocours." In ORPHEE-RDV 2017, Atelier "Méthodologies et outils pour le recueil, l'analyse et la visualisation de traces numériques d'interaction."
- Fuchs, B. (2017b). "Assister l'utilisateur à expliciter un modèle de trace avec l'analyse de concepts formels." 28es Journées francophones d'Ingénierie des Connaissances IC 2017, Caen (France), pp. 151-162.
- Fuchs, B. (2018). "Focaliser l'extraction d'épisodes séquentiels à partir de traces par le contexte." 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018, 6 juillet 2018, Nancy (France), pp. 213-227.
- Gilliot, J-M., Iksal, S., Medou, D., Dabbebi, I. (2018) "Conception participative de tableaux de bord d'apprentissage." IHM'18 : 30e Conférence Francophone sur l'Interaction Homme-Machine, Brest, France. Actes téléchargeables depuis le site web IHM-2018, p119
- Harrak, F., Bouchet, F., Luengo, V. (2017) "Identifying the Relationships Between Students' Questions Type and their Behavior." 10th International Conference on Educational Data Mining, EDM 2017. Best poster award.
- Harrak, F., Bouchet F., Luengo V. (2019) "From Students' Questions to Students' Profiles in a Blended Learning Environment." *Journal of Learning Analytics*. (1).
- Houssein, S., Di Marco, L., Schwebel, C., Luengo, V., Morand, P., Gillois, P., (2019) "Consequences of Switching to Blended Learning: The Grenoble Medical School Key Elements", *Studies in Health Technology and Informatics*, vol. 247, pp. 356-360, (IOS Press)
- Itani, A., Brisson, L., Garlatti, S., (2018) "Understanding Learner's Drop-Out in MOOCs", Yin Hujun; Camacho David; Novais Paulo; Antonio J; Tallón-Ballesteros. *Intelligent Data Engineering and Automated Learning -- IDEAL 2018*, Springer International Publishing, pp.233-244, Lecture Notes in Computer Science book series (LNCS, volume 11314), 978-3-030-03493-1. <10.1007/978-3-030-03493-1\_25>
- Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.. (2010) "A Data Repository for the EDM community" In *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press.
- Lebis, A., Lefevre, M., Luengo, V., Guin, N. (2016) "Towards a Capitalization of Processes Analyzing Learning Interaction Traces." 11th European Conference on Technology Enhanced Learning (EC-TEL 2016), Lyon, France. Springer, Adaptive and Adaptable Learning, Proceedings, 9891, pp.397-403, Lecture Notes in Computer Science.

- Lebis, A., Lefevre, M., Luengo, V., Guin, N. (2017) "Approche narrative des processus d'analyses de traces d'apprentissage : un framework ontologique pour la capitalisation." In N. Guin, B. De Lièvre, M. Trestini, & B. Coulibaly (Eds.), Actes de la 8ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (pp. 101- 112). Strasbourg: ATIEF.
- Lebis, A. (2018) "Assistance à la réutilisation de processus d'analyse de traces d'apprentissage via une approche narrative et sémantique." Septièmes Rencontres Jeunes Chercheurs en EIAH (RJC EIAH 2018) , Besançon, France. Actes des 7ièmes RJC-EIAH 2018.
- Lebis, A., Lefevre, M., Luengo, V., Guin, N. (2018) "Capitalisation of Analysis Processes : Enabling Reproducibility, Openness and Adaptability thanks to Narration." LAK '18 - 8th International Conference on Learning Analytics and Knowledge, Sydney, Australia. ACM, Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp.245-254
- Mandran, N., Ortega, M., Luengo, V., Bouhineau, D. (2015) "DOP8\_Cycle: Merging both data and analysis operators life cycles for Technology Enhanced Learning." Learning Analytics and Knowledge, Poughkeepsie, United States. <hal-01167195>
- May, M., Iksal, S. (2016) "The Side Effect of Learning Analytics: An Empirical Study on E-Learning Technologies and User Privacy." In Communications in Computer and Information Science, edited by Springer.
- May, M., Iksal, S., Usener, C. A. (2016) "Learning Tracking Data Analysis: How privacy issues affect student perception on E-learning?" 8th International Conference on Computer Supported Education (CSEDU), Rome, Italy, pp. 154-161.
- Mille, A., Peres-Labourdette Lembe, V. (2017) "Learning Analytics : vers une éthique par construction des EIAH." In ORPHEE-RDV 2017, Atelier "Méthodologies et outils pour le recueil, l'analyse et la visualisation de traces numériques d'interaction."
- Pierrot, L., El-Kechaï, H., Iksal, S., Cerisier, J.-F. (2017) "Etude de la circulation des pratiques numériques juvéniles : approche par la prescription de l'observation." In N. Guin, B. De Lièvre, M. Trestini, & B. Coulibaly (Eds.), Actes de la 8ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (pp. 77-88). Strasbourg: ATIEF.
- Reffay, C., Betbeder, M. and Chanier, T. (2012) "Multimodal learning and teaching corpora exchange: lessons learned in five years by the Mulce project". In *Int. Journal Technology Enhanced Learning*, Vol. 4, Nos. 1/2, pp.11–30.
- Reffay, C., May, M., Mille, A. (2017). "French HUBBLE project experience on Ethics in Learning Analytics." Workshop Ethics and Privacy for Learning Analytics in European Conference on Technology Enhance Learning (EC-TEL 2017), Tallinn, Estonie. HAL : edutice-01633343
- Romero, C. and Ventura, S. (2007) "Educational Data Mining: « A Survey from 1995 to 2005»". *Expert Systems with Applications*, 33(1), pp. 135-146.
- Siemens, G., Baker, R.S.J.d. (2012) "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration". *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*.
- Sanchez, E., Kalmpourtzis, G., Cazes, J., Berthoix, M., Monod- Ansaldo, R. (2015) "Learning with Tactileo Map: From Gamification to Ludicization of Fieldwork." *GI Forum journal, Geospatial Minds for Society*(1), 261-271.
- Sanchez, E., Emin Martinez, V., Mandran, N. (2015) "Jeu-game, jeu-play vers une modélisation du jeu. Une étude empirique à partir des traces numériques d'interaction du jeu Tamagocours." *STICEF*, 22(1). Sanchez, E. (2016) "Characterizing Social Engagement in a Digital Role-Playing