

## A LIRE AVANT DE RENSEIGNER LE DOCUMENT

Ce document servira à identifier les besoins en collecte et analyse de données pour la création d'indicateurs et à connaître les pratiques en analyse de données dans le cadre de Hubble. Cette synthèse permettra la création du modèle conceptuel pour la plateforme Hubble.

En fonction du travail de collecte et/ou d'analyse que vous avez fait, vous renseignerez uniquement les parties qui concernent votre travail. Nous vous demandons d'être le plus précis possible dans vos descriptions.

=====

**Date de rédaction** : 17 juillet 2016

**Nom du rédacteur du document** : Matthieu Cisel

Spécialités : Didactique, Sciences de l'éducation, Informatique, Analyse de données (*rayez les mentions inutiles*)

**Cas d'étude Hubble** : MOOC (MOOCAZ, Effectuation)

**Scénario hubble** : 2-4

**Personnes impliquées pour la collecte et l'analyse** : Matthieu Cisel, Sebastien Iksal

**Période de la collecte** : 11/2013-06/2015

**Période de l'analyse** : 01/2016 - 06/2016

=====

### **Dispositif d'apprentissage (Etude de cas de Hubble)**

**Type de dispositif** : MOOC

**Finalité de l'apprentissage** : selon le MOOC, apprendre à concevoir un MOOC (MOOCAZ), diriger une entreprise (Effectuation), conception et mise en oeuvre d'algorithmes (ALGO)

**Utilisation du dispositif et fonctionnalités** : Visionnage de vidéos, réalisation d'exercices automatisés, rendu de devoirs.

**Contexte de production de données** : Les données sont produites pendant des sessions de formation d'une durée s'étalant de cinq à sept semaines. Les sessions du MOOCAZ analysées ont été organisées en juin 2014 et en novembre 2014. Les sessions de

Effectuation analyse ont été organisées en novembre 2013, mai 2014 et novembre 2014. La session du MOOC ALGO analysée date de novembre 2013.

*(Au besoin indiquer les différents moments de la production (savoir si des données ont été produites sur plusieurs années)*

### **Décrire la problématique posée et les objectifs de l'analyse :**

1/ Nous nous intéressons aux modes de suivi du cours pour le MOOCAZ, un mode de suivi étant défini comme la combinaison de différentes actions (visionnage de vidéo, réalisation de quiz, rendu de devoirs) au sein d'un module du cours donné. Un mode de suivi alternatif correspond à un mode de suivi qui, s'il est adopté pour l'ensemble des modules d'un cours, ne permet pas l'obtention du certificat. La problématique est la suivante : quels sont les principaux modes de suivi alternatifs, que représentent-ils sur le plan quantitatif par rapport aux modes de suivi débouchant sur l'obtention du certificat ?

2/ Sur le MOOC Effectuation. On cherche à déterminer dans quelle mesure les certifiés d'un MOOC (participants qui obtiennent le certificat) négligent les vidéos du cours (au sens où ils ne les visionnent pas); c'est l'échantillonnage. On cherche à appréhender l'ampleur de l'échantillonnage et à identifier sur quels éléments de la séquence pédagogique il prévaut le plus.

### **Production des données :**

Décrire le processus de production des données brutes : Les données sont récoltées et anonymisées par les plates-formes hébergeant les cours (FUN pour MOOCAZ, Unow pour Effectuation, Coursera pour le MOOC d'algorithmique), puis transmises aux chercheurs.

Liste des variables initialement recueillies :

- visionnage de vidéos
- réalisation d'un exercice automatisé
- rendu d'un devoir
- obtention du certificat

Plateformes/outils utilisés:

**FUN :**

Points forts	Points faibles
	Disparition d'une partie importante des logs

Nombreux évènements indisponibles sur les autres plates-formes (interruption d'une vidéo, etc)	(entre 40% et 80% selon les cours)  Pas de données sur le téléchargement des vidéos pédagogiques
--	--

**Unow :**

Points forts	Points faibles
Complétude des données	Pas de données sur le téléchargement des vidéos pédagogiques

**Coursera :**

Points forts	Points faibles
Complétude des données. Données sur le téléchargement des vidéos pédagogiques. Organisation des données en tables SQL.	Pas d'identifiant commun dans les tables SQL entre actions réalisées dans les forums de discussion et autres types d'actions (visionnage de vidéos, etc).

**Stockage des données:**

Plateformes/outils utilisés: Serveur du laboratoire STEF

Points forts de ces plateformes	Points faibles
Pérennité des données	Les données ne sont accessibles que

	depuis le laboratoire
--	-----------------------

### Description des pré-traitements:

**Objectifs des pré-traitements** : Permettre l'analyse des données par les outils UTL ou R

Décrire le processus de pré-traitement : Seules les données issues de la plate-forme FUN ont fait l'objet d'un pré-traitement, les données issues de Unow ou de Coursera ayant déjà subi un pré-traitement par les plates-formes : construction d'une matrice ressources pédagogiques x utilisateur pour Unow, et organisation en table SQL pour Coursera. Les processus de pré-traitement mis en oeuvre par ces entreprises ne sont pas connus. En ce qui concerne les données de FUN, deux méthodes ont été employées, seule la dernière ayant été retenue. Les logs sont composés d'un mélange de JSON et d'autres formats de données. L'objet du pré-traitement est de décomposer les logs de sorte à ce que ces différents éléments constitutifs (nature de l'action réalisée, horodatage, etc.), soient compris dans des colonnes distinctes.

**Excel :**

Points forts	Points faibles
Grande rapidité de traitement des données et d'exécution, pas de développement sur mesure à réaliser	Limitations fortes relatives au pré-traitement des données brutes (format hybride JSON) lorsque celles-ci dépassent une certaine taille.

**UTL+moulinette de traitement des logs :**

Points forts	Points faibles
Décomposition systématique et contrôlée des logs bruts.	Développement sur mesure nécessaire, spécifique à la plate-forme FUN.

**Description des analyses :** (Faire une description de chacune des analyses conduites)

### Liste des variables :

Nom de la variable	Description
visionnage de vidéos	L'évènement apparaît lorsque l'utilisateur appuie sur le bouton lecture de la vidéo
réalisation d'un quiz	L'évènement apparaît lorsque l'utilisateur réalise un exercice automatisé, qcm, réponse numérique, etc
rendu d'un devoir évalué par les pairs	correspond au cas particulier des devoirs corrigés automatiquement par un autre participant
rendu d'un devoir corrigé par un programme	correspond au cas particulier des programmes informatiques corrigés automatiquement par un programme
obtention du certificat	Variable booléenne, qui prend la valeur 1 lorsque le participant obtient le certificat

### Liste des méthodes mise en œuvre :

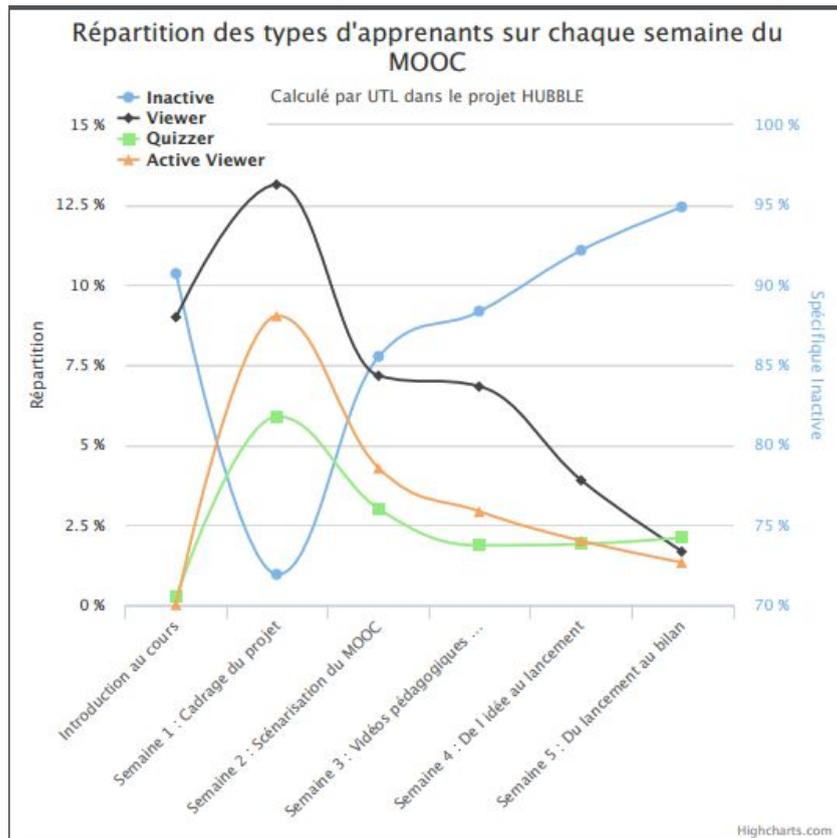
**Analyse 1** : Pour chaque module de cours et pour chaque individu, un mode de suivi est défini comme une combinaison d'actions. On visionne la succession et a l'importance relative des modes de suivi au fil des modules de cours.

**Analyse 2** : Pour Effectuation, pourcentage des ressources négligées par les certifiés, évolution au fil de la séquence pédagogique.

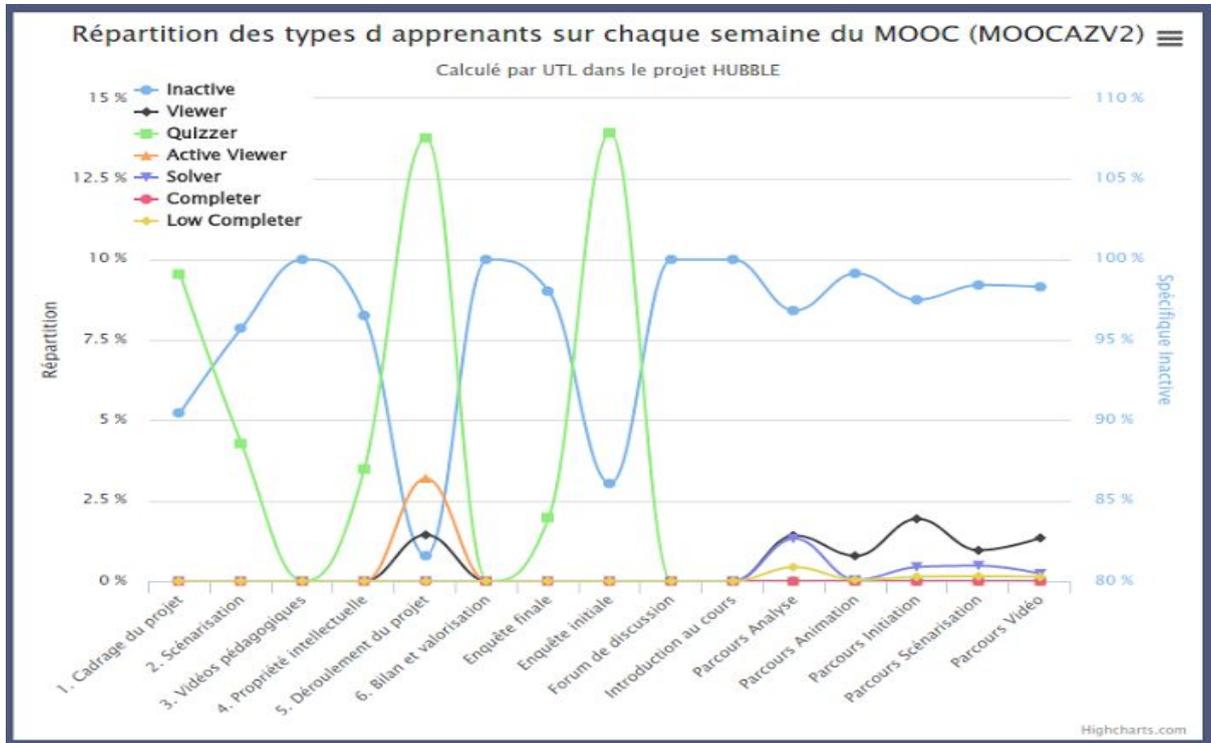
**Mode opératoire technique, logiciels utilisés** : Pour le MOOC ALGO, comme pour Effectuation, le programme R a été utilisé. Pour le MOOCAZ, l'analyse a été réalisée par la plate-forme UTL.

**Résultats obtenus** : Nombreuses interrogations quand au niveau de détail et d'explication a fournir ici (j'ai des pages de texte au besoin, mais ne mettrai que les graphiques)

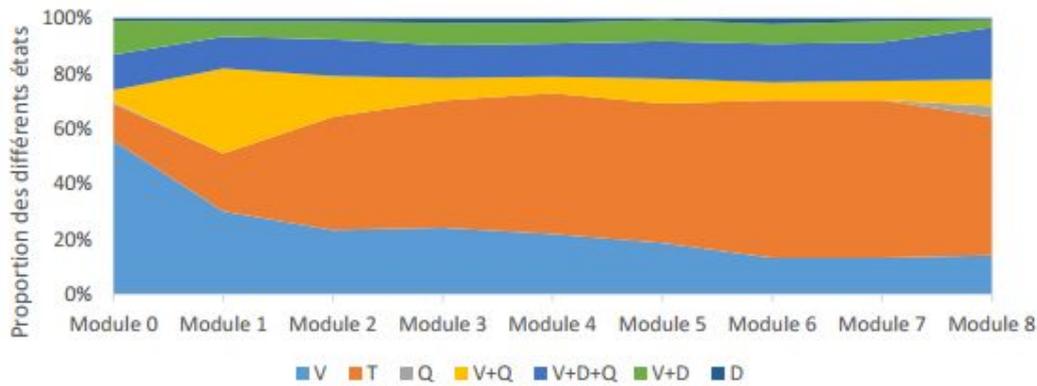
**Analyse 1** : MOOCAZ1, UTL



MOOCAZ2 :



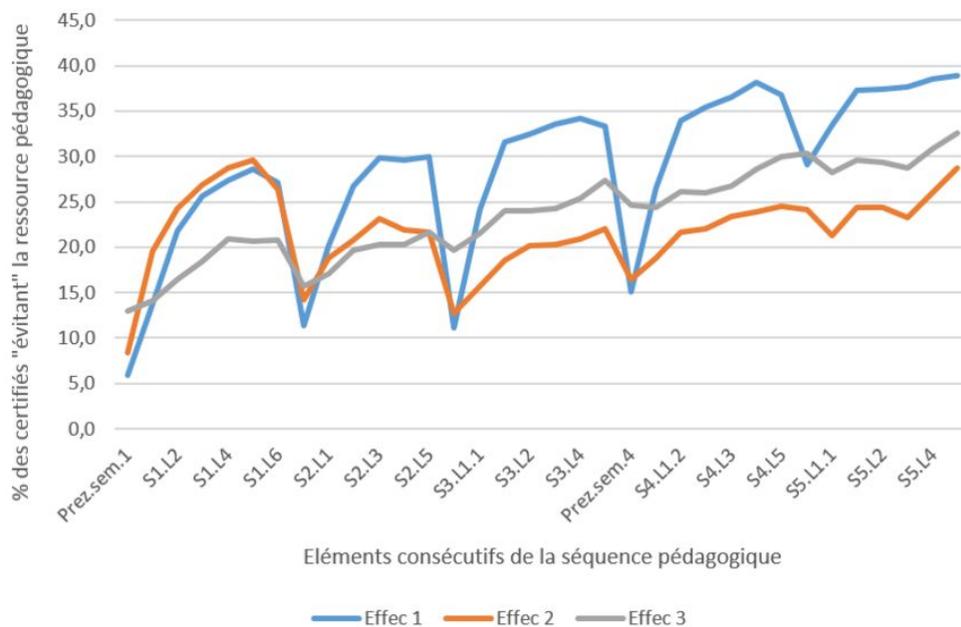
ALGO :



**Exemple de description :** On constate qu'à partir du module 3, le comportement de téléchargement domine parmi les participants actifs. Le visionnage en auditeur libre concerne près de 30% des inscrits au module 1, et sa proportion passe sous la barre des 15% pour le dernier module. La tendance est similaire pour le visionnage-rendu de quiz, dont la proportion est divisée par trois entre le module 1 et le module 2. Tout au long des modules du cours, l'importance du suivi assidu et quasi-assidu demeure relativement stable, autour de 10 et 5% des modes de suivi, respectivement. Le solvage ne représente quant à lui qu'une minorité des états ; sa proportion ne dépasse jamais 2% des inscrits actifs. Le téléchargement quant à lui, alors qu'il est minoritaire sur le premier module du cours, deviennent très rapidement majoritaire parmi les participants actifs.

**Analyse 2 :** Echantillonnage effectuation :

<https://github.com/hubble-learning-analytics/learning-analytics-catalog/tree/master/echantillonnage-au-sein-d-une-sequence-pedagogique>



Points forts des analyses	Points faibles des analyses
	Pour FUN, dans le MOOCAZ2, les modules comprenant les devoirs évalués par les pairs sont distincts de ceux qui comprennent les vidéos, ce qui change

### Description des données produites au cours du traitement

Objectif de la création de ces nouvelles données : Raisonner sur des comportements observables a l'échelle d'un module

Mode de calcul de ces variables : le mode de calcul est décrit dans la définition qui suit

### Description des nouvelles variables : Nom et description

Nom	description
visionnage en auditeur libre ou <i>viewer</i>	Ne faire que visionner des vidéos ( $v$ ), ou visionner et télécharger une vidéo ( $v+t$ ) est qualifié de visionnage en auditeur libre ( $V$ ), en référence à l'auditing de Kizilcec et al. (2013)
téléchargement	Le comportement qui consister à se cantonner à télécharger les vidéos ( $t$ ) d'un module est identifié comme téléchargement ( $T$ ).
solving, ou solver	Se contenter de réaliser des devoirs ( $d$ ) sera identifié comme du solving, par analogie avec le solver d'Anderson et al. (2014), et noté pour devoir.
Inactive	Si un participant ne laisse pas de trace d'activité, on parlera simplement d'inactivité. ( $I$ )
Active viewer ou visionnage-rendu de quiz	Nous parlerons de visionnage-rendu de quiz pour qualifier le comportement qui consiste à visionner au moins une vidéos, et réaliser au moins un quiz mais sans rendre de devoir ( $v+q$ et $v+t+q$ ), et nous utiliserons le dénominatif $V+Q$ pour identifier ce comportement
suivi assidu, ou	Visionner au moins une vidéo, soumettre au moins un quiz et

completer	rendre le devoir (v+q+d et v+q+t+d), sera considéré comme le suivi assidu; nous utiliserons le signe V+Q+D pour désigner ce comportement.
suivi quasi-assidu., ou low completer	Visionner des vidéos et rendre un devoir mais sans réaliser de quiz (V+D) est un comportement que nous identifierons comme le suivi quasi-assidu.

### **Description des Itérations**

Aucune itération n'a été réalisée