

## 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 : 27 juin 2016

Nom du rédacteur du document : Jean-François Cerisier, Hassina El Kechai, Laetitia Pierrot et Lucie Pottier

Spécialités : ~~Didactique, Sciences de l'éducation~~, Informatique, Analyse de données, Sciences de l'information et de la communication (*raier les mentions inutiles*)

Cas d'étude Hubble : TECHNE

Scénario hubble : S2, S3 et S4 (?)

Personnes impliquées pour la collecte et l'analyse : Jean-François Cerisier, Hassina El Kechai, Laetitia Pierrot, Lucie Pottier LIUM

Période de la collecte : 2015-2016

Période de l'analyse : juin-novembre 2016

=====

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

**Type de dispositif** : Accompagnement scientifique du projet Living Cloud qui fournit des équipements numériques nomades aux élèves du Lycée Pilote Innovant International (LP2I).

**Finalité de l'apprentissage** : Le projet Living Cloud a comme objectif de transformer les pratiques pédagogiques des enseignants ainsi que les conditions d'apprentissage des élèves. Le rôle du laboratoire Techné est alors d'observer l'impact du projet sur ces deux éléments à travers 5 sous-projets, dont le premier qui est basé principalement sur la collecte et l'analyse de traces. Ce sous-projet 1 a comme vocation de fournir des éléments de compréhension sur les usages numériques des lycéens.

**Utilisation du dispositif et fonctionnalités** : L'établissement étudié comporte 499 élèves en voie générale. Ces élèves ont été équipés dès leur rentrée en seconde d'une tablette ou ordinateur portable (ou hybride). Après diffusion d'une autorisation de participation au projet, 161 élèves (76 en seconde, 53 en première et 32 en terminale) ont accepté de fournir leurs données d'utilisation à partir du réseau du lycée (logs du proxy). De plus, 54 élèves ont accepté d'être équipés d'un logiciel de traçage (Kidlogger) permettant de suivre l'activité en ligne et hors ligne, à la fois dans l'établissement et également en dehors.

Sont tracés:

- l'équipement utilisé ;
- les sites, fichiers et dossiers consultés ;
- les applications lancées ;
- la durée et la période d'utilisation.

Ces diverses données ont vocation à alimenter un tableau de bord.

Afin de compléter l'analyse de ces données collectées automatiquement, des entretiens d'explicitation, individuels ou collectifs, sont organisés auprès des élèves volontaires.

**Contexte de production de données** : Les données du proxy sont disponibles sur la période de l'année scolaire, à savoir de septembre 2015 à juin 2016. Les logs du logiciel de traçage sont disponibles depuis avril-mai 2016. Une nouvelle session de collecte des données est en cours de préparation pour l'année scolaire 2016-2017.

*(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 :**

**Problématique générale** Comprendre les conditions d'appropriation à l'échelle collective du numérique et identifier les types de pratiques numériques des lycéens du LP21.

**Objectifs de l'analyse** : Les besoins d'analyse présentés ici sont liés au travail de recherche doctorale portant sur la circulation des pratiques numériques des lycéens. La circulation correspond dans notre cas au passage d'une pratique individuelle (par un lycéen) à une pratique collective (par plusieurs). La circulation renvoie donc à l'évolution des usages dans le temps. Le préalable à la compréhension de cette circulation est l'identification de pratiques. Les besoins d'analyse sont les suivants :

- identifier, établir une catégorisation des pratiques des lycéens, à partir des actions les plus fréquentes ;
- discriminer les actions en cours/hors cours ;
- discriminer les actions par classe (promo) ;
- discriminer les actions par genre ;
- situer dans le temps les actions d'un utilisateur ;
- identifier l'émergence d'une pratique individuelle à partir d'actions collectives (partagées) et discriminer les actions non partagées (isolées) ;
- identifier des groupes d'utilisateur ayant des actions communes ;
- vérifier si les communautés concordent avec les groupes prédéfinis ;

- vérifier s'il y a un lien entre l'organisation de la communauté (division du travail) et les actions réalisées.

### **Production des données :**

**Décrire le processus de production des données brutes :** Pour les logs du proxy, collecte systématique des sites consultés par utilisateur selon un login attribué (y compris Pop-up, ad, cookie...). Pour les logs de Kidlogger, collecte systématique des actions des utilisateurs par équipement tracé.

Extraction des données au format SQL à partir des bases de données hébergées à l'Université de Poitiers (Phpmyadmin)

### **Liste des variables initialement recueillies : Nom et Description**

#### **LOGS PROXY :**

- datetime indique la date et l'heure de connexion et est exprimé sous la forme JJ/MM/AAAA HH :MM.
- displayed\_login est l'identifiant attribué par l'établissement à l'élève, sous la forme NomInitialeprénom.
- ip correspond à l'adresse IP de l'équipement connecté.
- url correspond à l'adresse complète consultée.
- Domain précise le nom de domaine du site consulté.
- web\_protocol précise le protocole web utilisé.

#### **LOGS KIDLOGGER :**

- ID correspond à l'identifiant unique de l'action enregistrée
- Device\_id correspond à l'identifiant associé à l'élève tracé, suivant son équipement
- wcSection\_id et wcCategory\_id sont des colonnes à ignorer
- log\_type indique la nature de l'action. Il peut s'agir d'une application (app), d'une consultation web (url), d'une recherche en ligne (search-query), d'un appel entrant (in\_call) ou sortant (out\_call), d'une saisie de clavier (keystrokes), d'un dossier consulté (folder), d'une mise en veille (idle) ou d'un message entrant (in\_sms) ou sortant (out\_sms). Selon l'équipement tracé et l'installation retenue, les appels, messages et recherches ne sont pas tracés et n'apparaissent pas dans les logs.
- Duration précise la durée en secondes d'une action. Ce champ peut avoir pour valeur 0.
- Name qualifie l'action réalisée. En fonction de la classe, ce champ apporte des précisions sur l'action (nom de l'application ouverte ou contenu saisi au clavier ou url consultée).
- Date et time donnent des indications sur la période d'utilisation. Date est exprimé sur le format AAAA-MM-JJ et time HH :MM :SS.

Plateformes/outils utilisés: Kidlogger / base de données du proxy

--	--

Points forts	Points faibles

### Stockage des données:

Stockage sur une base de données :

- les logs du proxy sont hébergés sur un serveur de la Région Poitou-Charentes et transmis au laboratoire au format CSV. Ils sont ensuite stockés sur une base dédiée à l'Université de Poitiers et peuvent être réexportés au format SQL (CSV si besoin).
- les logs de Kidlogger sont hébergés sur une base de données l'Université de Poitiers et peuvent être exportés au format SQL (CSV si besoin).

Plateformes/outils utilisés: Base de données

Points forts de ces plateformes	Points faibles

### Description des pré-traitements:

Objectifs des pré-traitements :

- croiser les logs avec l'emploi du temps des élèves ;
- ajouter des informations complémentaires sur les utilisateurs.

Décrire le processus de pré-traitement :

- recodage des emplois du temps dans une table de données dédiée ;
- ajout des champs dans la table de logs de Kidlogger (jour, classe, niveau, genre, login du proxy, équipement, système d'exploitation).

Plateformes/outils utilisés:

Points forts	Points faibles

--	--

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

Liste des variables : Nom et Description

Liste des méthodes mise en œuvre :

Mode opératoire technique, logiciels utilisés

Résultats obtenus:

Points forts des analyses	Points faibles des analyses

**Description des données produites au cours du traitement**

Objectif de la création de ces nouvelles données :

Mode de calcul de ces variables :

**Description des nouvelles variables : Nom et description**

Nom	description

--	--

### Description des Itérations

Pourquoi le processus d'analyse a été itéré ?

Points forts des itérations	Points faibles des itérations