

GAIA-X Education & Skills - GT Traces d'Apprentissage

Réunion du **25 octobre 2021**

Participants :

- Matthias De Bièvre (Visions)
- François Bocquet (MENJS)
- Pierre Boulet (Université de Lille)
- Anne Boyer (LORIA, Université de Lorraine)
- Eric Cestari (Cstar Industries)
- Jean-Marie Gillot (IMT Atlantique)
- Quitterie Lucas (FUN)
- Julien Maupetit (FUN)
- Azim Roussanaly (LORIA, Université de Lorraine)
- Matt Sonnati (Inokufu)
- Sylvie Tissot (Tralalere)

La réunion commence par un bref tour de table des participants, puis une définition des traces et des événements d'apprentissage par le LORIA et FUN.

Le focus de cette première réunion porte sur l'interopérabilité des formats et des outils de collecte de traces d'apprentissage, point de départ nécessaire à la constitution/agrégation de jeux de données homogènes pour entraîner des algorithmes d'apprentissage automatiques (*machine learning*) .

Formats des traces d'apprentissage

Parmi l'ensemble des entités représentées, il semblerait qu'il y ait convergence sur le format utilisé pour représenter les traces d'apprentissage : **xAPI** standard est utilisé par le LORIA, Inokufu, Tralalère et FUN.

Dans le cadre de leur projet LOLA (Laboratoire Ouvert en Learning Analytics), le LORIA a mis en place des jeux de données pour la recherche en utilisant des profils xAPI [1]. Les profils xAPI peuvent être vus comme une commune sur laquelle nous grammons nous fonder pour décrire des événements d'apprentissage. Ce pourrait être un des livrables de ce groupe de travail (GT).

Remarques :

- Nous devons définir quels sont les types d'activités prioritaires qui donneront lieu à la mise en place d'un profil xAPI commun spécifié par ce GT.
- Une fois le projet LOLA rendu public, le LORIA pourrait nous présenter les profils sur lesquels ils ont travaillé comme base de discussion.
- Azim mentionne par ailleurs que dans le cadre de ce projet, ils ont pu mettre en place des procédures d'anonymisation des données. Ce point est à mettre à l'ordre du jour d'une prochaine session de travail.

Afin de définir nos priorités en terme de profil xAPI à spécifier, Matthias propose de décrire des cas d'usage concrets (*user stories*) dans un premier lieu en mettant l'accent sur ceux qui vont générer le plus de traction. De ces *user stories* vont découler à la fois la granularité des données à collecter et les bases légales nécessaires pour ce faire.

La question de la gouvernance autour de ces traces d'apprentissage est alors brièvement évoquée afin de définir un cadre de confiance dans lequel nous agissons. Ce GT pourrait organiser cette gouvernance (en association avec des utilisateurs de la plateforme).

Remarques :

- Le recueil du consentement doit-il être envisagé par plateforme ou sur une plateforme centralisée ?

- Des ateliers de sensibilisation aux données collectées doivent être organisées avec les utilisateurs des différentes plateformes qui organiseront la collecte (*data literacy*). Sylvie nous informe que Tralalère a une expertise en la matière.

Outils de collecte des traces d'apprentissage

L'Université de Lorraine ainsi que l'Université de Lille utilisent la suite d'outils *open source* Apereo Learning Analytics Initiative qui a été récemment renommé Shuhari [2]. Initialement, cette solution a été mise en place dans le cadre du projet Dune-Eole [3, 4].

Par ailleurs, dans le cadre du projet LOLA et suite au projet METAL, le LORIA après un benchmark des LRS (*Learning Record Store*) open source existants a retenu le projet Trax LRS [5] pour ses performances et continuent depuis à développer des modules spécifiques aux besoins de leur projet.

De son côté, Tralalère utilise Learning Locker [6], autre LRS open source populaire. FUN a aussi essayé ce dernier mais a finalement abandonné le projet, n'arrivant pas à gérer la mise à l'échelle du produit pour le volume de données générées par leur plateforme.

FUN et Inokufu n'utilisent pas des LRS à proprement parler mais des solutions de gestion de logs performantes dans le cloud (LDP/Graylog [7] dans le cas de FUN). Couplé à un datalake basé sur elasticsearch, le stockage des traces d'apprentissage offre des performances satisfaisantes pour mettre en place des tableaux de bords basés sur l'outil open source Grafana [8]. FUN travaille actuellement sur la couche de conformité à la spécification d'un LRS (API HTTP).

En conclusion, il apparaît que l'interopérabilité concernant les outils de collecte d'apprentissage soit garantie par la spécification du LRS [9]. Il convient cependant d'ajouter une couche de validation des événements soumis étant donné les profils xAPI définis dans la section précédente ; une grande attention devra être apportée à la gestion du contexte xAPI.

Remarques :

- Le format CMI5 [10] pourrait faire l'objet d'une étude poussée pour son intérêt dans la description de profils d'activités liées à un LMS.
- Comment gérer l'interop avec les formats propriétaires (moodle, edx) et les traces issues de systèmes basés sur SCORM ?
- Suite à une question d'Eric : il pourrait être pertinent de publier un annuaire de LRS acceptant de recevoir/envoyer des données depuis/vers des outils tiers lorsque les auteurs des outils ne sont pas en capacité d'héberger un tel service.

François Bocquet, en déplacement, avait des difficultés de connexion. Il nous transmettra ses remarques par écrit.

Feuille de route

Lors de notre prochaine réunion, nous devons brainstormer sur les premières user stories qui semblent les plus pertinentes en terme d'usage. De ces user stories vont découler des indicateurs qui pourront être mis en évidence par des profils xAPI. Avoir un triplet user story / indicateurs / profils xAPI complet serait un point d'étape de ce GT avant d'"industrialiser" la procédure.

Le sujet de l'anonymisation/pseudonymisation est transverse et doit faire l'objet d'une session de travail dédiée en association avec un (ou des ?) DPOs. Avoir un retour d'expérience du projet LOLA sur ce sujet est précieux et serait un bon point de départ.

La date de notre prochaine réunion n'a pas encore été fixée, FUN va proposer framadate pour Novembre.

References

- [1] profils xAPI <https://github.com/adlnet/xapi-profiles/blob/master/xapi-profiles-about.md>
- [2] Shuhari <https://www.apereo.org/projects/shuhari-new>
- [3] <https://factuel.univ-lorraine.fr/node/13920>
- [4] Dune-Eole <https://www.dune-eole.fr>
- [5] Trax LRS <http://traxlrs.com>
- [6] Learning Locker <https://learningpool.com/solutions/learning-record-store-learning-locker-draft/learning-locker-community-overview/>
- [7] OVH LDP (Log Data Platform) <https://www.ovhcloud.com/fr/data-platforms/logs/>
- [8] Grafana <https://grafana.com/>
- [9] LRS <https://xapi.com/learning-record-store/>
- [10] CMI5 <https://xapi.com/cmi5/>