

\*LUE IMPACT OLKi - Open Language and Knowledge for Citizens

\*Compte-rendu de la réunion #8 du groupe de travail sur la *plateforme OLKi*  
27 mai 2020 de 13h30 à 15h30, visioconférence

## Documents de travail

Dossier B'UL: <https://bul.univ-lorraine.fr/index.php/s/K5gWr1HljdofvZ>

Liens vers les cas d'usage :

<http://etherpadlite.univ-lorraine.fr/p/n9ksvzdbcx>

<https://olki.loria.fr/blog/usecase/>

## Compte-rendu

### 1) Etat d'avancement du développement de la plateforme

Release 0.4 beta : <https://framagit.org/synalp/olki/olki/-/releases>

Possibilité d'installer cette version beta et faire des retours de bugs à Pierre-Antoine. Pierre et Nicolas vont tenter l'installation.

La méthodologie Agile est mentionnée. Remarque : à chaque commit, le code est déployé sur le serveur.

### Fonctionnalités opérationnelles

- Abonnement entre serveurs
- Fédération des corpus (et des métadonnées) et des commentaires

Autre fonctionnalité initialement évoquée mais non implémentée à ce jour :

- Traitements sur les corpus : statistiques, visualisation d'occurrences de mots (voyant tools), création d'un formulaire, etc... → priorisation à investiguer (via un stage)

Protocole 3iF d'annotation des images: intéressant pour l'archivage, ciblé pour les images. N'est peut-être pas pertinent pour la plateforme.

Web annotation: <https://www.w3.org/ns/oa>

Faut-il être un démonstrateur assez riche ou se concentrer sur les fonctionnalités cœur ?

Faut-il limiter à des corpus textuels ? (car type scan nécessite un plus gros investissement)

A trancher, mais des avis sont plutôt de se focaliser sur les fonctionnalités cœur.

Importance d'avoir des utilisateurs pour faire décoller la plateforme. Pour avoir des utilisateurs, il faut des fonctionnalités non trouvées par ailleurs, par exemple les interactions autour de corpus dans la communauté des historiens des sciences.

Cibler une communauté ou plusieurs communautés ?

Rappel sur les corpus : tous les types de documents (texte, image, son)

Format acceptés : texte, annoté (avec balise TEI ou étiquette par exemple) ou non

Quid des ressources déjà annotées ?

Question sur la granularité des annotations et le nombre de formats supportés.

Réponse faite sur les limitations en termes de :

- sélecteurs disponibles pour chaque type de document
- visualisation via des outils déjà développés : lecteur pour chaque type de document

Quels corpus publics pourraient être utilisés ?

- corpus Henri Poincaré, par exemple que les lettres (Pierre)
- + corpus de Mathieu
- + cf liste initiale du projet

Perspectives pour les derniers mois de développement (jusque fin août pour l'instant):

- Annotation
- Page d'accueil avec guide aux utilisateurs
- Débogage

## 2) Projet de soumission d'un article

Objectif de soumettre à une issue spéciale du journal Semantic Web Journal : (<http://www.semantic-web-journal.net/blog/call-papers-special-issue-semantic-technologies-data-and-algorithmic-governance>)

Article rédigé en partie seulement, une relecture sera demandée prochainement:

<http://etherpadlite.univ-lorraine.fr/p/I58vfy90op>

Attention : il y a un risque que l'article ne soit pas assez centré sur le web sémantique, mais la décision du journal d'accepter ou pas l'article dépendra aussi des autres soumissions.

Remarque : la plateforme n'a pas de technologie triplestore.

## 3) Propositions de sujet de stage L3 MIASHS (8 semaines)

Développement d'un système d'annotation web de documents textuels : Pas de candidat

Formalisation des besoins des utilisateurs de la plateforme OLKi : Chloé en stage du 01/06/2020 au 24/07/2020 devrait préciser les cas d'usage.

## 4) Interaction avec des réseaux

- Infrastructure européenne CLARIN

La plateforme pourrait être un service qui serait proposé via cette infrastructure, mais a priori CLARIN n'offrira pas de financements directs pour cela. Christophe est en contact.

- Infrastructure de recherche européenne DARIAH

Working Group de DARIAH pourrait être une bonne solution pour donner de la visibilité à la plateforme. Adeline Joffres est membre du JRC de DARIAH qui gère et évalue les Working Groups. Elle a été contacté, sans réponse.

- Consortium CORLI

Le consortium CORLI, financé par Huma-Num, pourrait apporter un support. Christophe Parisse est en charge du consortium. Christophe Benzitoun des ATILF est impliqué également. Contact pas encore pris.

## 5) Futur de la plateforme

Quelles sources de financement?

- Appel FMSH : <http://www.fmsch.fr/fr/recherche/30863>

Non répondu car ne couvre pas les frais de personnel et par manque de temps

- Projet H2020 ?
- Appel sur infrastructure

- <https://socialhub.activitypub.rocks/t/grants-and-funding-for-our-projects/86>
- Fonctionnalités payantes

## 6) Divers

L'UL est en réflexion pour développer un entrepôt pour les données de recherche (Dataverse) : 2 instances logicielles pour l'instant

Phase de test avec plusieurs labo pilotes (IJL, BETA, ...) pour sélectionner des jeux de données, etc...et pour choisir une solution

Cas d'usage éventuel avec la plateforme OLKi: proposer un workflow direct pour la soumission de données de la plateforme OLKi vers l'entrepôt de stockage

### **Prochaine rencontre**

Proposition : Jeudi 18 juin 14h

Rédactrice : Aurore Coince

Pour rejoindre la visioconférence

- connectez-vous sur <https://visio.inria.fr> (<login>@inria.fr) puis "Se réunir" : gt.plateforme@visio.inria.fr
- pour les externes <https://visio.inria.fr/invited.sf?secret=7OJkn78ZSd2mT0hjm0gBOg&id=130784964>
- depuis un terminal de visioconférence : appelez visio.inria.fr (IP: 128.93.162.102) puis entrez le numéro de la conférence : 130784964#
- par téléphone : +33 4 92 38 77 88 (77788), puis entrez 130784964#