

EzPAARSE-ezMESURE : mettre à profit les Open Badges pour favoriser le travail collaboratif et produire de meilleures analyses

Thomas Porquet

Département Services et Prospective - Consortium Couperin
23 rue Daviel
75013 Paris

Dominique Lechaudel

Inist – CNRS
Équipe Mesure des Usages
2, rue Jean Zay
54519 Vandoeuvre-lès-Nancy

Résumé

Les abonnements aux revues et ouvrages scientifiques sont très coûteux : les budgets pour permettre aux étudiants, aux enseignants et aux chercheurs d’y accéder s’élèvent à environ 100 millions d’euros par an pour l’ESR français. Les outils d’évaluation de leur usage sont stratégiques, autant pour la politique documentaire que budgétaire. Ces problématiques sont partagées par tous les établissements.

Depuis 2013, le logiciel libre ezPAARSE permet à chaque établissement d’analyser les traces de passage des usagers vers ces ressources. Spécialisé dans l’identification des ressources consultées, ezPAARSE se nourrit d’un travail hautement collaboratif, qui a lieu sur la plateforme Analogist où la structure de chaque plateforme éditeur est décortiquée pour déterminer la sémantique des URL. Ce travail (déclarer une nouvelle plateforme, ajouter une analyse, implémenter le parseur correspondant, etc.) est bien délimité et, depuis peu de temps, validé et récompensé par la remise d’Open Badges, standard sécurisé défini par la fondation Mozilla, un système de médailles virtuelles à exposer sur AnalogIST ou des réseaux sociaux comme LinkedIn, Twitter ou Facebook. De cette façon, nous valorisons et motivons la participation de collègues répartis dans un grand nombre d’établissements, qu’ils soient en France ou à l’étranger.

Le dispositif est désormais complété par ezMESURE, l’entrepôt national de données. Basé sur la suite Elastic Search + Kibana il permet de visualiser sous forme de “facettes” et d’agrégations, les données chargées depuis les instances locales d’ezPAARSE. Cet environnement est proposé par un partenariat fructueux entre l’Inist-CNRS et le consortium Couperin.org.

Mots-clefs

Données d’usage, ressources électroniques, abonnements, analyse de log, tableaux de bord, open badge, travail collaboratif, COUNTER

0 Note préliminaire

Cet article, dans son contenu et sa structure, est très proche de l'article publié par les mêmes auteurs pour la session S15 - *Data intelligence in libraries: the actual and artificial perspectives* - Big Data, de la conférence satellite de l'IFLA qui s'est tenue à Francfort en août 2019 [1].

Dans le présent article, l'accent est placé sur la partie collaborative du dispositif et l'organisation qui en découle. Les éléments techniques et les outils décrits sont quant à eux en tout point identiques.

1 Introduction

Dans cet article, nous nous intéresserons aux outils à l'instar d'Open Badge et aux méthodes qui nous aident à mutualiser, à rendre visible et à valoriser des tâches collaboratives qui enrichissent substantiellement les outils que nous proposons.

Pour bien comprendre le contexte et les enjeux dans lesquels nous nous trouvons, nous décrivons d'abord de quoi sont composés les différents éléments qui constituent la chaîne de traitement et d'analyse, formée par ezPAARSE¹ (décrit en partie 2.2) et ezMESURE² (partie 2.3), et mise à disposition de la communauté (établissements membres du consortium Couperin.org en particulier). Cette chaîne permet la génération de statistiques d'utilisation des ressources numériques par les établissements qui s'y abonnent, mais indépendantes de celles que les éditeurs³ peuvent leur fournir. Procurer un niveau de détail que n'offrent pas les statistiques des éditeurs, et proposer un dispositif qui permet le croisement de ces données pour plusieurs établissements sont les objectifs qui nous animent, car ils sont jugés stratégiquement importants.

La capacité d'ezPAARSE à reconnaître, à analyser et à qualifier finement les consultations dépend des parseurs⁴ avec lesquels il est livré, chacun étant spécialisé pour une plateforme d'éditeur distincte (partie 4,2). Ces parseurs sont le résultat de la collaboration entre les documentalistes des établissements qui analysent les plateformes éditeurs⁵ sur le site public AnalogIST (partie 4,1)⁶ et les informaticiens du projet qui prennent le relais pour les implémenter et les verser sur la forge logicielle GitHub. Ces échanges et cette collaboration sont rendus explicites et valorisés par la mise en place des Open Badges.

Les badges ezPAARSE matérialisent donc, selon des critères d'attribution que nous détaillerons plus loin, l'implication dans une communauté et l'engagement dans l'amélioration d'un l'outil open source. Leur interopérabilité permet aux membres de la communauté de faire valoir ces nouvelles compétences auprès de leurs collègues, sur les réseaux sociaux comme LinkedIn, Twitter ou Facebook.

1. <https://www.ezpaarse.org/>

2. <https://ezmeasure.couperin.org/>

3. Nous utilisons le terme « éditeurs » pour parler de tous les opérateurs qui fournissent et commercialisent des contenus numériques (sous forme de revues, bases de données, ouvrages numérique, etc.) auxquels s'abonnent les établissements pour leurs étudiants, chercheurs, enseignants, personnels techniques et administratifs.

4. La description détaillée de ce qu'est un parseur se trouve dans la section 3.1. Vous pouvez voir un parseur comme un petit bout de programme qui sait reconnaître et extraire des éléments précis dans une URL, pour leur faire correspondre des informations sur le type de ressource qui est consultée.

5. Une plateforme éditeur est à comprendre au sens COUNTER (voir l'entrée « platform » du glossaire <https://www.projectcounter.org/appendix-glossary-terms/>) : « Interface d'un agrégateur, d'un éditeur ou d'un autre service en ligne qui fournit du contenu à un utilisateur. »

6. <http://analyses.ezpaarse.org>

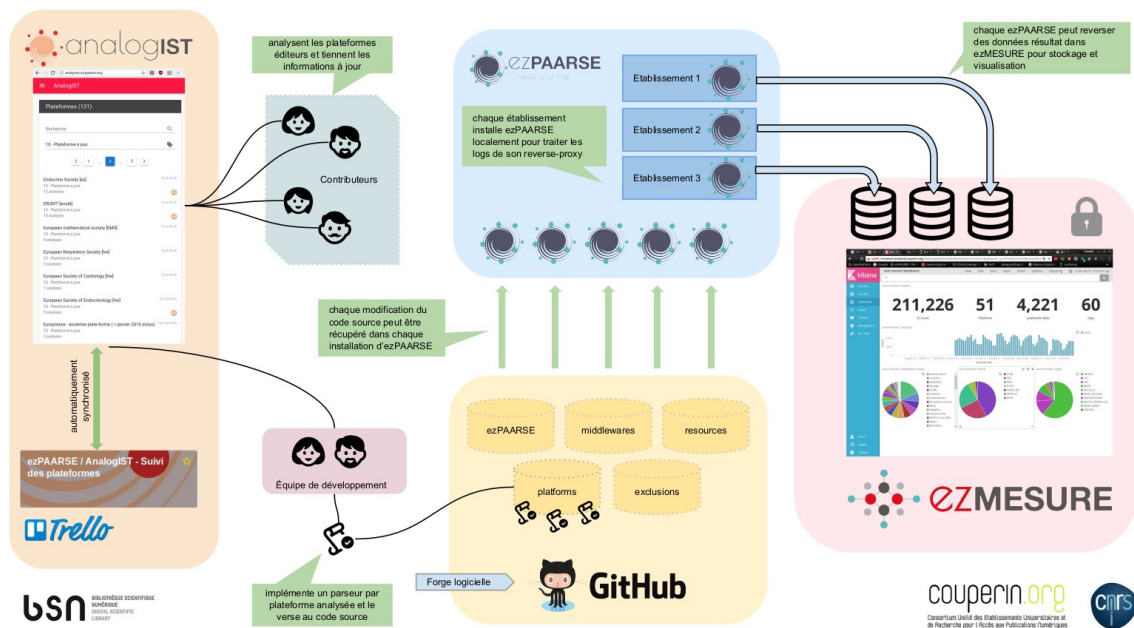


Figure 1 - Ecosystème AnalogIST – ezPAARSE – ezMESURE : vue d’ensemble

2 ezPAARSE et ezMESURE : principes de fonctionnement

2.1 L’analyse de logs

L’analyse des fichiers logs est une tâche technique relativement classique. En témoigne la grande variété de logiciels d’analyse web disponible dans le domaine : Awstats, Piwik (devenu Matomo), etc. Ce principe (des logs et un outil qui les analyse pour déterminer une audience) est aussi celui qui est généralement mis en œuvre par les fournisseurs dans le domaine des statistiques d’utilisation, lorsqu’il s’agit d’identifier et rendre compte des accès réussis ou refusés aux ressources électroniques par les établissements de l’enseignement supérieur et de la recherche qui s’y abonnent.

Et, tout comme les éditeurs de plateformes peuvent analyser les logs de leurs serveurs, les établissements et institutions abonnés qui utilisent un serveur proxy pour authentifier leurs utilisateurs sont aussi en capacité d’analyser le trafic sortant pour en extraire des éléments à grain fin et que nous nommons « événements de consultation » dans la suite de l’article.

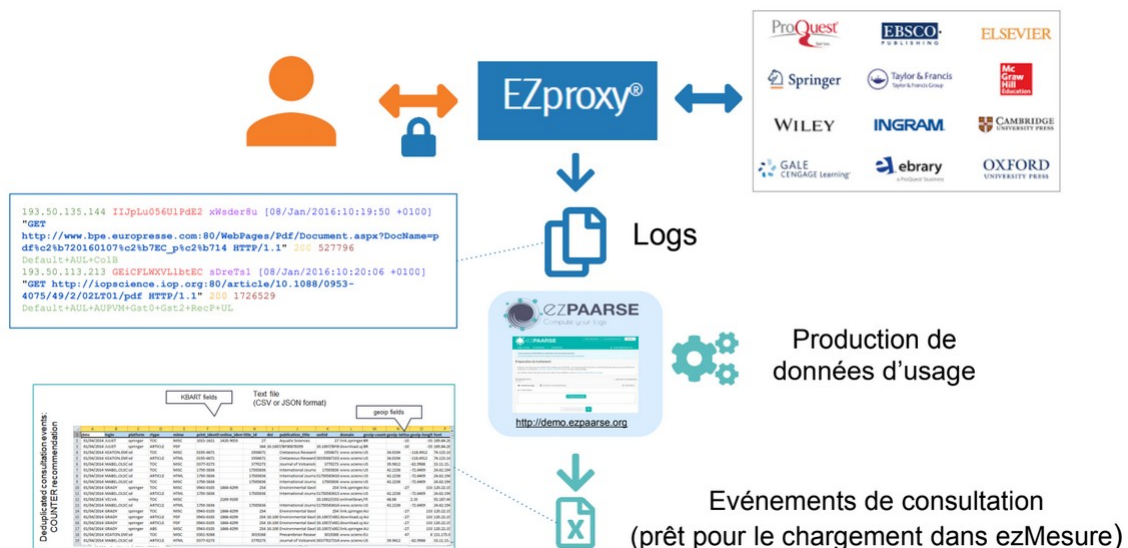


Figure 2 - EzPAARSE : production d'événements de consultation à partir de lignes de logs d'un proxy

2.2 Traiter des logs avec ezPAARSE

2.2.1 Une spécialisation aux ressources documentaires

Le logiciel libre et gratuit ezPAARSE est disponible pour la communauté depuis 2013[2]. Fruit d'une collaboration nouée de longue date entre le consortium de négociation documentaire Couperin.org, l'INIST (unité propre de service du CNRS et plusieurs établissements pilotes, dont l'Université de Lorraine. Son intérêt réside dans :

- le filtrage de tout ce qui n'est pas jugé pertinent (et qui compose en fait l'écrasante majorité des logs) ;
- l'enrichissement des événements de consultation retenus avec des données bibliographiques explicites à chaque fois que cela est possible.

D'installation aisée et rapide, ezPAARSE s'utilise ponctuellement ou, mieux, s'automatise pour être déclenché à fréquence régulière pour traiter les logs à mesure qu'ils sont produits. La mise en place et l'automatisation incombent généralement au service informatique de l'établissement intéressé.

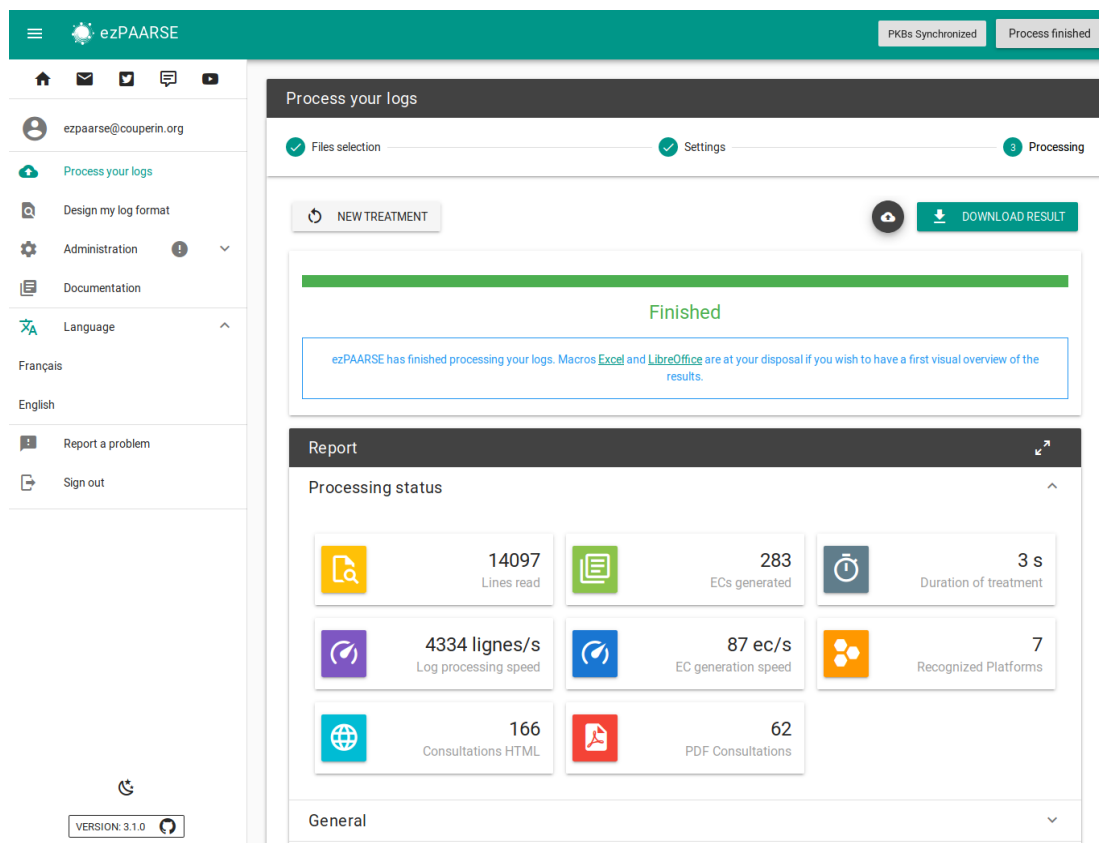


Figure 3 - un fichier logs correctement analysé par ezPAARSE

2.2.2 Construire sur les expériences du passé

Avant la création d'ezPAARSE, l'INIST-CNRS utilisait un système similaire, NumStats, développé et utilisé en interne uniquement. La maintenance de cette application désormais remise, représentait une charge de travail importante et difficilement soutenable pour l'unique personne qui en avait alors la charge. Les résultats obtenus, jugés extrêmement intéressants par un grand nombre de collègues d'autres établissements, prouvaient que le mécanisme d'analyse des logs mis en œuvre était valide et opérationnel. Les résultats obtenus étaient d'une grande finesse, au prix toutefois d'une maintenance de tous les instants pour tenir les capacités de reconnaissance à jour.

Techniquement repris à zéro avec des outils jugés plus adaptés (Node.JS plutôt que PHP, MongoDB plutôt que MySQL), ezPAARSE est mené en partenariat entre l'INIST-CNRS, l'Université de Lorraine et le consortium Couperin.org, implémenté sous licence libre et proposé gratuitement dès le début des développements.

La qualité des résultats fournis par ezPAARSE provient presque directement d'un fort travail collaboratif pour étendre et tenir à jour ses capacités de reconnaissance des ressources consultées sur les différentes plateformes web des éditeurs. Ce travail-là, identifié comme essentiel dès les premières itérations de sa conception, est découplé du cycle de développement du cœur applicatif d'ezPAARSE et pris en charge dans un environnement distinct, baptisé AnalogIST, décrit plus bas.

2.3 Visualiser ses données d'usage dans des tableaux de bord dynamiques

2.3.1 La production de données comme étape préalable à leur visualisation

Pouvoir produire des données d'usage n'est que la première étape, désormais bien identifiée par les établissements utilisateurs.

La deuxième étape consiste ensuite à visualiser et analyser les données produites par les différentes instances d'ezPAARSE installées localement dans les établissements. ezMESURE entre ici en jeu : référentiel national d'indexation et de stockage, les établissements y trouvent aussi un outil de conception et de visualisation de tableaux de bord.

2.3.2 L'entrepôt national ezMESURE

Depuis le début de l'année 2016, l'INIST-CNRS, Couperin.org et ses membres ont poursuivi le travail entamé avec la conception d'ezPAARSE en créant un entrepôt basé sur la suite ElasticSearch et Kibana : ezMESURE. Il héberge les données d'accès de près de 60 établissements et leur permet de les visualiser à la demande et sous forme de tableaux de bord et de visualisations personnalisables.

Il est à noter qu'un établissement qui dépose ses données ezPAARSE dans ezMESURE est le seul à y accéder : son espace de stockage lui est réservé et les données qui s'affichent dans ses tableaux de bord ne sont visibles que par lui. C'est le principe qui s'applique par défaut.

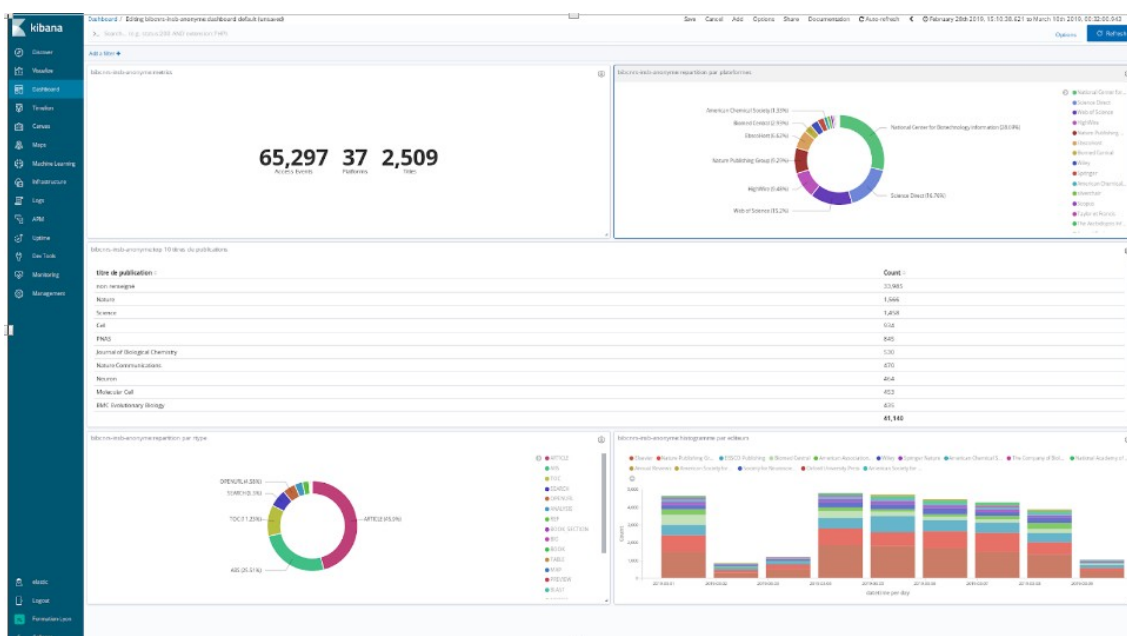


Figure 4 - Un tableau de bord Kibana dans ezMESURE

Avec cet écosystème, nous fournissons à nos utilisateurs une chaîne de traitement et d'exploitation complète, pour qu'ils reprennent le contrôle sur des données stratégiques.

3 L'intérêt des analyses de plateformes

3.1 Une analyse sur AnalogIST, ça consiste en quoi ?

La plateforme AnalogIST est l'endroit où, pour chaque plateforme d'éditeur ou de fournisseur, les différents types d'URL sont collectés, découpés semi-automatiquement et commentés si nécessaire, pour que le parseur qui en sera tiré puisse être à la fois le plus précis et complet possible [3]. Pour les éditeurs qui ne sont pas encore reconnus par ezPAARSE ou qui ont simplement besoin d'une mise à jour, il est facile de venir contribuer à une analyse de plateforme : pour cela, il suffit d'ouvrir un compte Trello et de suivre le guide⁷.

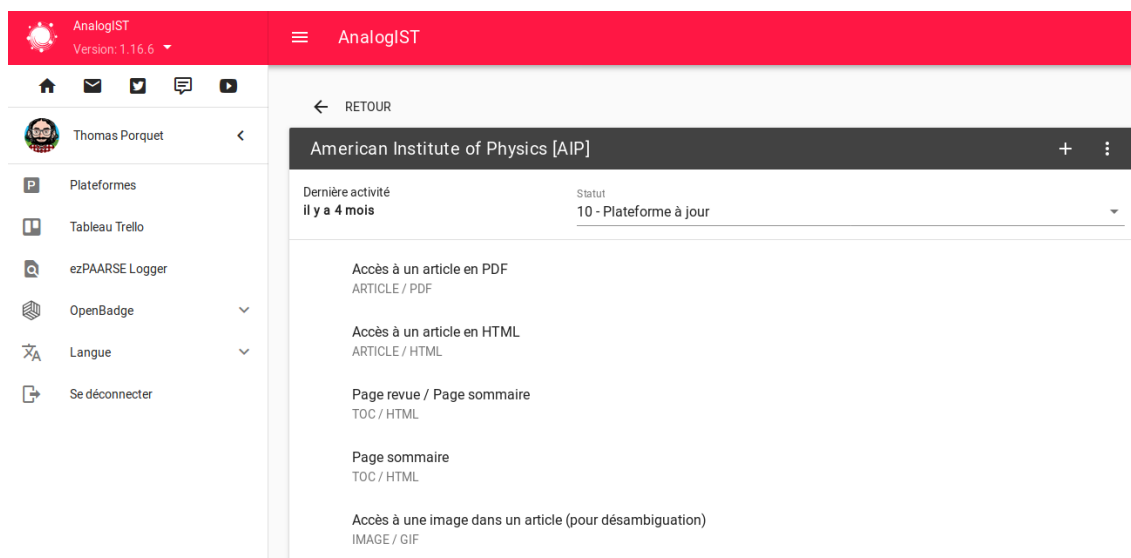


Figure 5 - L'interface AnalogIST permettant d'éditer les analyses de plateformes

3.1 Les parseurs, résultats formalisés des analyses

Une fois l'analyse jugée suffisante, ou complète, un informaticien (généralement de l'équipe ezPAARSE/ezMESURE) prend le relais pour implémenter le parseur correspondant.

Les parseurs constituent le cœur du mécanisme de reconnaissance de la sémantique des URL par ezPAARSE. Un parseur est un morceau de code, spécifiquement dédié aux types d'URL d'une plateforme éditeur, traduction formalisée et programmatique de l'analyse réalisée par un documentaliste. Ce dernier relève, par simple copier-coller sur AnalogIST, les URL des types d'objets documentaires consultables par ses utilisateurs et dont il souhaite mesurer l'usage.

Le parseur est déclenché par le nom de domaine depuis lequel la ressource est servie. Ce nom de qui lui-même indique la plateforme éditeur depuis laquelle est servie la ressource.

Le parseur transforme une **URL** en un **événement de consultation** en distinguant la ressource unique consultée grâce à un identifiant de titre de revue ou d'article ainsi que son type MIME (PDF, HTML, etc.).

7. <https://blog.ezpaarse.org/2017/11/ateliers-et-formations-supports-et-guides/>

L'architecture d'ezPAARSE est prévue pour pouvoir intégrer facilement de nouveaux parseurs par l'utilisation d'un dépôt GitHub spécialisé et un mécanisme de mise à jour automatisable embarquée dans le logiciel.

3.2 Les parseurs : partagés et disponibles pour tous

Un nouveau parseur vient s'ajouter aux parseurs existants dans le dépôt GitHub dédié⁸ et peut être récupéré immédiatement par toutes les instances d'ezPAARSE par un mécanisme simple de mise à jour embarqué dans la zone d'administration du logiciel. Ce mode de fonctionnement garantit que chaque nouveau parseur et chaque mise à jour du logiciel et de ses ressources profitent à toutes les installations existantes d'ezPAARSE : le travail n'est nécessaire qu'en un endroit et une seule fois.

3.3 Vérifier le bon fonctionnement d'un parseur

Enfin, pour vérifier simplement que les parseurs livrés se comportent correctement, une extension de navigateur, ezLOGGER⁹, est disponible. Elle capture momentanément le trafic d'un parcours sur une plateforme et l'envoie vers l'instance de démonstration d'ezPAARSE qui traite en direct les traces de passage et génère des événements de consultation pour peu que le ou les parseurs correspondants se soient correctement déclenchés.

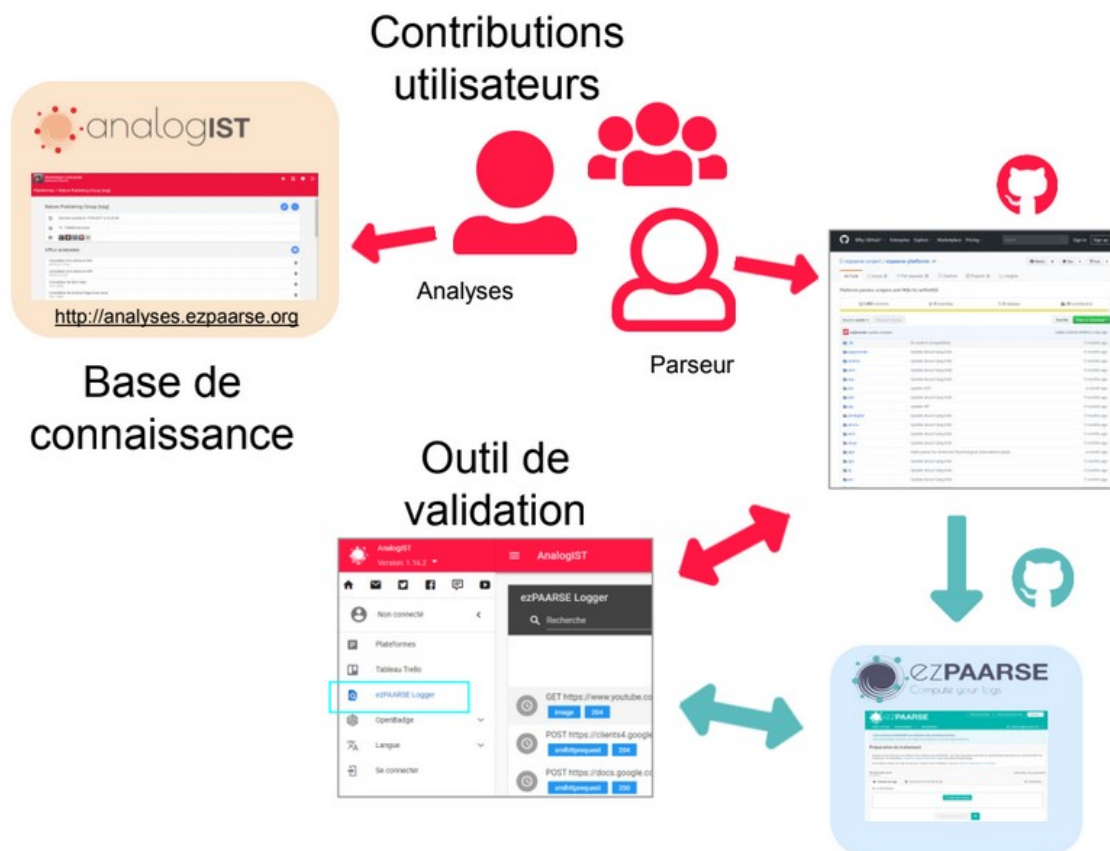


Figure 6 - AnalogIST : des analyses de plateformes aux parseurs

8. <https://github.com/ezparse-project/ezparse-platforms/>

9. <https://blog.ezparse.org/2017/11/tutoriel-ezlogger-fait-peau-neuve-dans-analogist/>

3.4 Les contributions au code source

La mise à disposition d'ezPAARSE sous la forme d'un logiciel Open Source permet d'accueillir les contributions. Au cours de la vie du logiciel une seule contribution, substantielle, a concerné la partie cœur, mais les contributions sous forme de « Pull Requests » (terminologie Git) pour la reconnaissance de nouvelles plateformes sont régulières et donnent lieu à des remises de badges.

4 Organiser et faire vivre le travail collaboratif

La collaboration aux activités décrites dans cet article sont généralement d'une nature très ponctuelle : analyser une plateforme ne prend par exemple qu'une heure ou deux, si on débute ; la routine de vérifier qu'un parseur continue d'être actuel n'en prend qu'une dizaine. Le partage de ces tâches prend alors tout son sens : dès lors qu'il est réparti correctement et reconduit régulièrement, ce travail assure à ezPAARSE une qualité reconnue largement. C'est avant tout pour encourager la prise en main de l'écosystème (outils, procédures) que les badges nous sont utiles : récompenser les premiers efforts, et fidéliser la continuité par la progression dans les types de badges (bronze, argent puis or).

Mais les badges n'existent pas tout seuls : ils ne sont que la partie la plus visible d'un ensemble d'actions et de procédures mises en place préalablement ou en parallèle : mode agile, rendez-vous en ligne, supports personnalisés ou en groupe, en présentiel ou à distance. C'est la variété des efforts déployés qui nous permettent aussi de fédérer une communauté^[4] dont les participants sont par ailleurs pris par des myriades d'autres tâches.

4.1 Les Open Badges : standard, forge et remise

L'« Open Badge » est un standard proposé par la fondation Mozilla qui le définit comme une sorte de médaille virtuelle, liée à une personne et qui peut contenir des informations sécurisées.

Un OpenBadge ressemble à une image mais encapsule des métadonnées sécurisées dont l'encodage est normé. L'administration des badges est une activité qui peut être déléguée. Dans le cas des badges ezPAARSE et ezMESURE, c'est le service OpenBadge Factory qui a été choisi pour la forge des badges (c'est le rôle « issuers » de la figure 6). Les badges, une fois acceptés par les utilisateurs, peuvent être visualisés (rôle « displayers » de cette même figure).

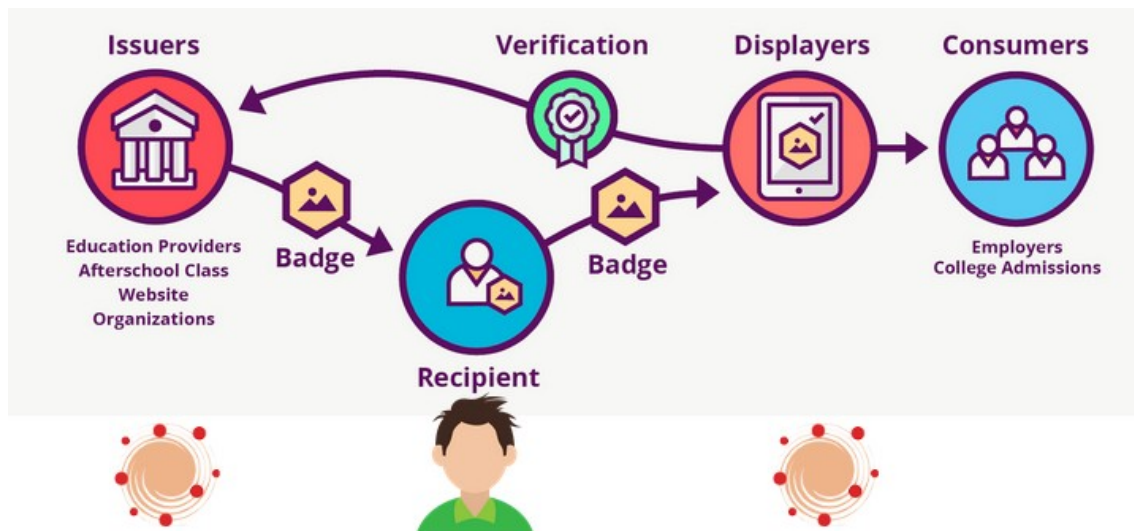


Figure 7 - Les acteurs de la chaîne d'attribution de badges et les interventions d'AnalogIST aux niveaux de l'attribution et de la visualisation (Source : The Open Badges Ecosystem, by Erik Knutson, Concentric Sky, licensed CC-BY)

Un badge ezPAARSE obtenu peut être exposé sur OpenBadge Passport¹⁰ ou AnalogIST mais aussi partagé sur des réseaux sociaux tiers : LinkedIn, Twitter, Facebook et motiver la participation d'autres collègues. La participation de la communauté est ainsi encouragée et récompensée par des badges ouverts qui en valident les différentes étapes : déclaration d'une nouvelle plateforme à prendre en compte, ajout d'une analyse, création ou mise à jour du parseur correspondant.

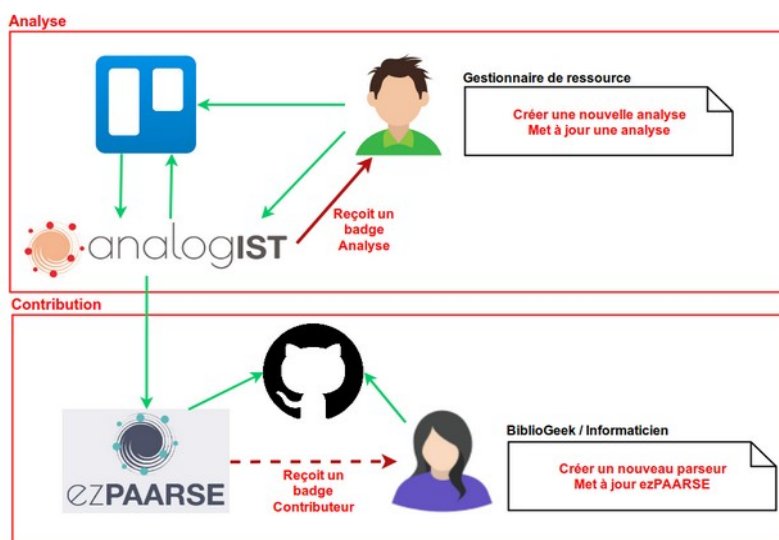


Figure 8 - Les deux types de contributions récompensées par un badge ezPAARSE

10. <https://openbadgepassport.com/> et un tutoriel pour y valoriser ses badges <https://blog.ezpaarse.org/2019/02/faq-comment-valoriser-mes-open-badges-ezpaarse-dans-open-badge-passport/>

Outre l'aspect ludique des badges, ce sont de réelles compétences qui sont identifiées dans les profils de poste des collaborateurs et utilisateurs de ces outils. Après un peu plus d'un an de fonctionnement, 75 badges pour les analyses de plateformes et les créations de parseurs ont été délivrés par notre serveur de badges.

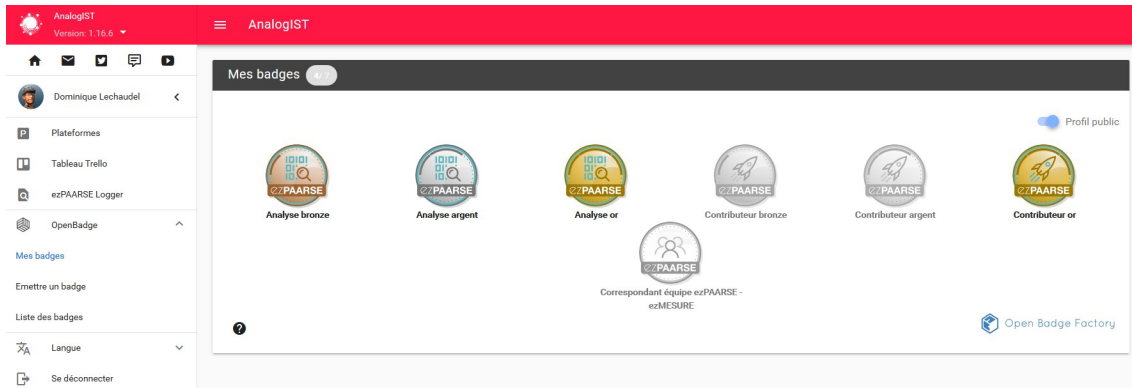


Figure 9 - Les différents badges ezPAARSE

En plus des badges ezPAARSE, il existe aussi des badges dédiés à ezMESURE. Chaque établissement déposant désigne deux correspondants :

- un correspondant “technique” qui est le garant du bon fonctionnement de la suite logicielle en local ;
- un correspondant “documentaire” qui est chargé d’évaluer la couverture documentaire de son établissement telle que la reflète ezPAARSE, de surveiller l’apparition de nouveaux parseurs ou leurs mises à jour et de gérer les accès aux données déposées dans ezMESURE pour les autres membres de son établissement.

A ce jour, 94 badges de correspondants ont été attribués, qui matérialisent là aussi des activités spécifiques en lien avec le projet global.

4.2 Les rendez-vous et les formations

L'équipe ezPAARSE propose deux formes d'assistance pour la mise en place de la suite ezPAARSE-ezMESURE :

- une journée de formation en présentiel et découpée en deux parties, l'une consacrée à ezPAARSE et l'autre à l'usage d'ezMESURE et la création de tableaux de bords ;
- des “rendez-vous ezPAARSE/ezMESURE” en ligne, d'une heure environ pour les utilisateurs d'un établissement qui souhaitent faire un point d'étape ou bénéficier d'une assistance à la mise en place d'une des briques de la chaîne : installation d'ezPAARSE sur un serveur local ;
 - automatisation des traitements ;
 - automatisation des chargements dans ezMESURE ;
 - prise en main des tableaux de bord Kibana.

Au cours d'un sprint typique (itération de 6 à 8 semaines), nous consacrons entre 5 et 20 heures au support.

4.3 La communication sur les réseaux

L'accompagnement et la formation sont une grande part de l'activité menée autour de l'écosystème décrit, qui se matérialise par différents supports accessibles aux utilisateurs :

- des partages d'écran enregistrés (ou « screencasts ») pour détailler les étapes techniques (installation, utilisation, automatisation) sur YouTube¹¹ ;
- des articles explicatifs dans le blog¹² ;
- des séances de formations thématiques en présentiel ou par partage d'écran. Des échanges (peu fréquents) sur les listes de diffusion ezpaarse-users et ezmesure-users¹³

Le blog rassemble une mine d'informations : des tutoriels techniques ou non, une rubrique FAQ, l'annonce des ateliers et les principales actions de communication. Toutes les mises à jour des plateformes y sont aussi consignées. Conçu en mode participatif modéré, la communauté peut l'enrichir et documenter leurs cas d'usage.

Enfin, la présence sur Twitter¹⁴ permet une communication directe avec la communauté pour répondre rapidement à des questions ponctuelles, rappeler des rendez-vous importants ou annoncer de nouvelles "fonctionnalités phares".



Figure 10 - Un utilisateur expose la réception de son badge sur son profil Twitter

4.4 Retour d'expérience

Après la mise en place initiale et plutôt exploratoire d'un serveur de badges à la faveur d'un stage de licence professionnelle au printemps 2018 puis son intégration complète à la plateforme AnalogIST dans les mois qui ont suivi, nous proposons quelques éléments de réflexion.

Passées les premières annonces, présentations et remises de badges, nous n'avons pas le sentiment d'avoir rencontré un engouement tout à fait à la hauteur de ce que nous aurions pu espérer. Aucun collègue ne nous demande par exemple spontanément comment obtenir un badge, ou comment progresser pour obtenir un meilleur badge. Ce premier constat n'est toutefois pas celui d'un échec. D'abord parce que le

11. <https://www.youtube.com/ezpaarse>

12. <https://blog.ezpaarse.org/>

13. <https://groupes.renater.fr/sympa/info/ezpaarse-users> et <https://groupes.renater.fr/sympa/info/ezmesure-users>

14. <https://twitter.com/ezpaarse>

dispositif des badges n'a pas été pensé comme la pierre angulaire de notre dispositif d'animation de communauté. Nous l'envisageons plutôt comme un « plus » ou un bonus, d'autant plus intéressant qu'il est largement automatisé et ne demande pas de maintenance particulière. Et ensuite parce que le dispositif même de badge n'est pas encore vraiment installé dans le monde des bibliothèques académique et universitaires. Il nous faudrait conduire une étude plus approfondie des systèmes de badges déployés ici et là pour pouvoir approfondir cet aspect en particulier [5].

De notre point de vue, les badges sont d'une grande aide pour aider nos utilisateurs à identifier explicitement les endroits où leurs participations sont possibles et attendus. En particulier, ils nous permettent de valoriser les actions de formation au moment où elles ont lieu : lors d'un atelier, chaque participant se voit remettre un premier badge à sa première analyse de plateforme.

Notre outillage pédagogique déployé autour de la suite logicielle s'en trouve complété. Une corde de plus à notre arc, en somme.

5 Conclusion

Installer ezPAARSE aujourd'hui, c'est bénéficier d'une liste de près de 250 parseurs, résultat d'un travail d'analyse décentralisé fourni par la communauté. Les parseurs assurent aux établissements utilisateurs une bonne couverture, en progression constante, de leurs abonnements aux ressources électroniques. Depuis l'été 2018, différentes actions, dont celles d'analyser des plateformes et d'écrire des parseurs, sont récompensées par la remise de badges. Ce dispositif, qui permet d'identifier et de valoriser un travail jusqu'alors largement souterrain, vient s'ajouter aux outils et méthodes d'animation de communauté déjà en place.

Les perspectives ouvertes par cette modalité d'animation nous permettent d'envisager des actions à mener, plus ciblées, autour de la mise à jour des analyses ou de la maintenance des parseurs, sous la forme par exemple de badges « spéciaux ».

Bibliographie

- [1] Porquet T. et Lechaudel D.. *ezPAARSE and ezMASURE: Assembling dashboards on a national repository from fine-grained and locally generated usage statistics to electronic resources*. IFLA, Francfort, Août 2019 <http://library.ifla.org/2722/1/s15-2019-porquet-en.pdf>
- [2] Porquet T., Lechaudel D., Gully S., & Jouneau T.. Le dispositif ezPAARSE/AnalogIST : pour une analyse mutualisée des logs d'accès aux ressources documentaires payantes. JRES 2013 https://2013.jres.org/archives/33/paper33_article.pdf
- [3] Lechaudel D., Porquet T., Fabry C., Gully S., Jouneau T. & Schurter Y.. AnalogIST / ezPAARSE: analysing locally gathered logfiles to determine users' accesses to subscribed e-resources. In LIBER 43rd Annual Conference. Riga, Lettonie.
- [4] Ribas S., Guillaud P., Ubeda S.. *Logiciels & objets libres : Animer une communauté autour d'un projet ouvert*. Framasoft (coll.Framabook), 2016.
- [5] Ravet S., *Réflexions sur la genèse des Open Badges, Distances et médiations des savoirs*, décembre 2017, <http://journals.openedition.org/dms/2043>