

Rapport de mission : Séjour au CLEO (Centre pour l'Édition électronique Ouverte, Marseille)

Dates : 19 et 20 avril 2016

Participants : B.-O. DOZO, I. MAYEUR, F. PROVENZANO

Cadre de la mission

La mission a été réalisée dans le cadre de l'ARC « Genèse et actualités des humanités critiques » (GENACH), coordonnée par F. PROVENZANO et G. CORMANN à l'Université de Liège; et concerne plus spécifiquement le projet de thèse d'I. MAYEUR, doctorante sur bourse NON-FRIA, dirigée par B.-O DOZO et F. PROVENZANO.

La thèse d'I. MAYEUR porte sur les conditions de production et de circulation des savoirs au sein des plateformes numériques en sciences humaines, et s'inscrit dans l'axe « Humanités numériques » du projet ARC. Le volet principal de son corpus est constitué des revues et carnets hébergées par le portail OpenEdition, géré par le CLEO (Centre pour l'Édition électronique ouverte), dont le siège est basé à Marseille. Plusieurs contacts ont été pris avec les responsables de ce portail, qui ont souhaité une rencontre pour établir avec eux les modalités d'exploitation scientifique de leurs données. I. MAYEUR sera amenée à fréquenter durablement les responsables d'OpenEdition dans le cadre de sa thèse ; raison pour laquelle la présence des co-promoteurs était indispensable, afin de délimiter le cadre scientifique du projet de thèse et poser des bases de travail optimales pour la suite.

Déroulement de la visite

Mardi 19 avril

Prise de contact

B.-O. DOZO, F. PROVENZANO et I. MAYEUR arrivent à 13h au CLEO, pour un lunch avec M. DACOS, directeur, et E. FAATH, chargée de projet. La prise de contact se poursuit entre 14h et 15h30 par la visite des locaux du CLEO et la rencontre avec les équipes d'OpenEdition.

Une **première réunion de travail** se tient ensuite entre 15h30 et 17h (participants : M. DACOS, B.-O. DOZO, E. FAATH, I. MAYEUR et F. PROVENZANO).

Présentation de l'environnement technique d'OpenEdition

M. DACOS débute en présentant l'environnement d'OpenEdition, à partir du rapport d'activité 2015. Ce rapport contient notamment un schéma d'information permettant d'aborder les questions liées à l'accès aux données. E. FAATH poursuit en présentant la structure [OpenEditionLab](#), destinée à promouvoir des pratiques innovantes impulsées par des recherches sur OpenEdition. Les activités du Lab se déploient dans trois secteurs :

- **Recherche** : permet de faire progresser les frontières de la connaissance dans divers domaines. Par exemple, en informatique, il s'agira d'augmenter le pourcentage de résultat d'un algorithme déjà connu (par exemple, l'algorithme de détection des références bibliographiques);

- **Développement** de nouvelles fonctionnalités sur base des recherches (par exemple, le logiciel Bilbo à partir des recherches précitées) ;
- **Innovation** : établir une passerelle entre l'invention et l'innovation ; cette dernière étant le succès de l'invention. Au cœur de l'innovation, se trouve la transmission, qui permet d'introduire de nouveaux usages, des pratiques inédites.

Les intérêts de chacune des parties représentées diffèrent. Les questions de recherche intéresseront davantage les chercheurs que les membres d'OpenEdition. Par contre, OpenEdition sera attentif à ce qu'il est possible d'extraire en termes de développement ou d'innovation à partir de la recherche. À cet égard, l'étude récente du sociologue Julien Bordier (2016) portant sur les pratiques d'*Open Commentary* et *Open Peer Reviewing* développées sur le carnet de la revue *VertigO* constitue une recherche porteuse de recommandations à même de faire évoluer le dispositif de la plateforme.

M. DACOS poursuit en présentant l'architecture système de la plateforme OpenEdition, et l'organisation des 25 serveurs qui la composent, à partir du rapport d'activité 2015 toujours. Le portail se démarque par le grand nombre de comptes en écriture qui s'y inscrivent (11000, tandis que d'autres plateformes n'en accueillent qu'une douzaine). Un contingent important d'utilisateurs peut ainsi modifier des pages. Le système information de la plateforme est ensuite détaillé. L'ensemble des données de la plateforme est aplati dans CORE, ce qui signifie que tous les liens existant entre les champs, les tables et les enregistrements sont supprimés et qu'il ne reste plus dès lors que les données brutes.

Il s'agit donc de déterminer quel type de données sera nécessaire pour la constitution du corpus de thèse d'I. MAYEUR ; sachant par exemple que les commentaires laissés sur les carnets de recherche *Hypothèses* ne figurent pas dans CORE, et que les logs de connexion n'apparaissent pas dans le système d'information. B.-O. DOZO suggère qu'I. MAYEUR se familiarise dans un premier temps avec les différents outils de production et de recherche dans les données, et détermine ensuite ce qui devra être affiné en fonction des besoins.

Des statistiques d'utilisation sont disponibles à l'adresse usages.openedition.org, qui croise des données issues d'awstat.org avec des données provenant de lodel. Ce dispositif permet la consultation de données relatives à un article, une revue, un auteur... Un outil existe également pour les logs de connexion sur l'interface : logs.openedition.org, qui recense les affluents du site déjà calculés par awstat. Les données sont exportables en Excel. Il convient toutefois de se montrer prudent dans l'interprétation de celles-ci, en raison du grand nombre de bruits présent. Ainsi, les anomalies les plus saillantes sont probablement d'origine non humaine (*bots*). L'outil devrait également disposer d'un résolveur de lien. C'est à cet endroit qu'on pourrait voir le trafic interne entre carnets et revues.

L'observation de **Twitter** permet d'étudier la diffusion des contenus de la plateforme. À cette fin, E. FAATH a soigneusement collecté les citations d'OpenEdition sur le média social, à partir des résultats provenant de l'API de Twitter qu'elle a ensuite exportés sur Excel et nettoyés, avant d'établir dans le tableur des statistiques sur cette base. Les quatre plateformes totalisent quelques 15000 tweets par mois. L'avantage de l'observation de Twitter est qu'il est possible de faire une enquête sur le profil des utilisateurs, et, par-là, de voir comment les communautés intègrent des éléments non scientifiques, ou comment se déplacent les frontières disciplinaires.

M. DACOS souligne que les commentaires laissés sur la plateforme OpenEdition ne constituent que la partie émergée de l'iceberg, le reste du champ étant occupé par ce qu'il nomme la « conversation silencieuse » : tweets et autres échos dans les médias sociaux, rétroliens et fréquentation. Il serait ainsi possible de suivre un auteur sur la plateforme, de voir ce qu'il a consulté, produit par la suite... Un projet d'identifiant unique pour les auteurs en SHS est d'ailleurs en chantier. Il s'agirait alors de reconstituer une **capsule conversationnelle** autour d'un objet textuel, soit les tweets, logs,... La

fonction « cité par... » sera bientôt disponible sur *Revue.org*. Également en développement, le projet HIRMOS (High Infrastructure Research Monographs), financé au niveau européen, vise à implémenter une dizaine de services dans six plateformes européennes, dont OpenEdition. De nouveaux usages devraient se développer ensuite.

F. PROVENZANO s'enquiert des balises exploitables qualitativement dans le corpus, pour explorer la question des genres textuels en usage au sein de la plateforme. M. DACOS indique qu'il existe sur *Revue.org* une typologie des documents ; par contre, *Hypothèses* ne reconnaît que le genre « billet », qui est massivement employé par les usagers des carnets de recherche. Seule une fraction minime des utilisateurs, peut-être 1%, connaît le moyen de catégoriser autrement les documents : dès lors, se fonder sur ce type de catégorisation des billets reviendrait à mesurer un degré de littéracie numérique. B.-O. DOZO considère quant à lui que les deux approches ne sont pas forcément indissociables.

De ce fait, les CR qui sont publiés sur la plateforme *Hypothèses* sont identifiés comme billets et non comme des recensions (car la catégorie créée par les usagers pour préciser le type de billet n'est pas reconnu comme métadonnée). Un projet en cours vise à repérer, dans les billets, ceux qui correspondent à des recensions, de manière à les relier éventuellement aux ouvrages publiés. L'algorithme développé par OpenEditionLab, fondé sur l'analyse textuelle, permet à ce stade d'identifier un billet comme CR avec 97% d'exactitude (reviewbooks.openeditionlab.org – accès restreint). E. FAATH signale à ce propos les travaux de Ch. BENKOUSSAS et de H. HAMDAM (thèses en informatique), dont il est fait mention sur le carnet d'OpenEditionLab (FAATH et al. 2013). Elle communiquera également les références de quelques études réalisées à partir de l'observation de la circulation des savoirs sur la plateforme (BRIATTE 2015, DEBATZ 2010, FAATH s.d., PUSCHMANN et BASTOS 2015).

Présentation du projet de thèse d'I. MAYEUR

La doctorante détaille ensuite son projet de thèse, ainsi que l'horizon de questionnement envisagé pour l'interrogation du corpus. La thèse s'inscrit dans le projet de l'ARC GENACH, qui étudie la circulation des savoirs dans les revues françaises et allemandes de l'après-guerre (jusqu'aux années 1980) ; et plus précisément dans l'axe consacré aux humanités numériques. Il s'agit ainsi de rendre compte des conditions de production, de diffusion et de réception des savoirs en contexte numérique, avec pour étude de cas les contenus de la plateforme OpenEdition, dont seront exploités les contenus textuels laissés par les utilisateurs: (hyper)textes, commentaires, échanges sur les médias sociaux, etc.

En résumé, la question principale est : comment (dans quel contexte matériel et textuel, sous quels formats, par quelles modalités) circulent les savoirs sur les plateformes d'édition numérique en SHS, et plus particulièrement OpenEdition ? Quels sont les usages qui s'y développent ? De quelle manière les modes de communication scientifique se voient-ils influencés par ces usages ?

Plusieurs axes de questionnement ont été dégagés à ce jour, devant être ajustés au corpus et à l'approche méthodologique:

- **Communautés actives sur les plateformes :**
 - o Qui sont-ils (chercheurs, amateurs, professionnels, groupes de recherche, etc.) et qu'y font-ils ? Existe-t-il une opposition pertinente entre usagers actifs (comité de revues, animateurs de blogs...) vs. passifs et, si oui, en quelle proportion ? De quelle manière s'organisent/se recoupent les pratiques d'écriture/lecture en ligne sur la plateforme OpenEdition ? Il ne s'agit pas ici de réaliser une étude sociologique, mais de déterminer les contours des communautés actives sur la plateforme et d'éventuels usages différenciés à partir de leurs productions discursives ;

- Comment ces usagers gèrent-ils leur *éthos* numérique, qu'il soit collectif, savant, profane (AMOSSY 2010)? Quelle scénographie/scène d'énonciation (MAINGUENEAU 2013) supporte les interactions des utilisateurs ou est créée par elles ?
- Quelles sont les pratiques d'appropriation des contenus sur la plateforme : parcours d'utilisateur, pratiques de citation (COMPAGNON 1979), moissonnage bibliographique, production et partage d'une littérature grise, etc. ? Comment s'exercent les fonctions critiques au sein de la plateforme (ex. : recensions, open commentary, open peer review...) (MAGIS et GRANJON 2016) ? Ce point peut notamment être étudié par le prisme des discours sur les pratiques : notices de présentation des revues/carnets ; posts de blogs jouant un rôle analogue : manifestes, etc. ;
- Peut-on dégager des logiques de fonctionnement communautaires au sein de la plateforme, articulées autour d'une revue/d'un carnet ?
- Quelles sont les autorités qui régissent les échanges sur la plateforme ? Comment s'exercent-elles dans l'économie d'une plateforme numérique SHS comme OpenEdition ?
- Quels échos de l'imaginaire social lié au monde de la recherche trouve-t-on au sein des textes de la plateforme (ex. : carnet de recherche reformaté en blog, comme manière de renouer avec la discussion de séminaire, universalisation de la recherche etc.) (DACOS et MOUNIER 2010) ? De quelle manière cet imaginaire permet-il l'anticipation des pratiques des usagers, et de quelle manière cette anticipation se marque-t-elle au sein de la plateforme (ex. fonction complémentaire des blogs pour l'actualité vs. revue vécue comme statique) (DAVALLON et al. 2003 ; SAEMMER 2015)?
- **Genres et formats textuels de la plateforme :**
 - Que contraignent et que permettent les cadres/architextes de la production scientifique numérique (articles, blogs et littérature grises) ? (GENETTE 1979 ; SOUCHIER 1998 ; DAVALLON et al. 2003 ; JEANNERET et SOUCHIER 2005 ; JEANNERET 2014) ;
 - En quoi ces formes sont-elles productrices de sens (MCKENZIE & CHARTIER 1991) ? Quels usages fait-on du paratexte (GENETTE 1979, 1982 ; 1987), de l'hypertexte (SAEMMER 2015)?
 - Quelles sont les différentes formes de textualité numérique en usage sur la plateforme (PAVEAU 2015) ? En quoi ces formes influencent-elles la pratique de la publication scientifique ? Y a-t-il des formes-modèles, des genres propres aux disciplines scientifiques ?
 - En quoi la forme « revue scientifique en SHS » se voit-elle affectée par ce changement de support (BERTHOU 2009) ? Avec quelles conséquences (flux, interactions, actualités, développement de nouveaux supports périphériques comme le carnet d'accompagnement...) ?;
- **Temporalités de la plateforme :**
 - Quelles sont les pratiques d'archivages en interne ? Cette archive a-t-elle vocation à être mise en dépôt, réutilisée ? À quelles fins ?
 - Comment se font les liens aux dépôts d'archives externes (ex. : HAL via Hypothèses) ? Avec quelle éditorialisation (épi-revues...) (Vitali-ROSATI 2016)?
 - Comment s'articulent les strates temporelles sur la plateforme, et en particulier le clivage archive/actualité (flux Calenda intégré sur Hypothèses et Revues.org ; actualités des revues...) ? De quelle manière les usagers s'approprient-ils ces flux ?
 - De quelle manière circulent ces flux (fragmentation, agrégation // pratiques de citation ?) (DACOS 2016) ;

- **Interdisciplinarité et circulation des concepts :**

- Comment la plateforme gère-t-elle les catégorisations disciplinaires de ses contenus ? D'où proviennent/comment sont élaborées ces classifications ? Comment ces catégorisations sont-elles appropriées par les utilisateurs (BONACCORSI et LABELLE 2010 ; CORDONNIER 2015)?
- Comment les concepts circulent-ils entre les disciplines (à travers les revues ou les blogs) ?

Sur la base de cet horizon de questionnement, il s'agira de prélever, au sein de l'immense corpus que constituent les contenus textuels d'OpenEdition, des portions relevantes pour dégager des réponses permettant de mettre en lumière les **logiques de circulation et d'appréhension des savoirs sur la plateforme** : pratiques d'éditorialisation par les usagers, formes d'appropriation dont elles témoignent, modes de diffusion des productions scientifique et fonctions critiques à l'œuvre en contexte numérique.

Mercredi 20 avril

B.-O. DOZO, F. PROVENZANO et I. MAYEUR arrivent à 9h30 au CLEO, pour une **deuxième réunion** avec les représentants du pôle informatique d'OpenEdition : Jean-François RIVIÈRE et Matthieu ORBAN. E. FAATH y assiste également.

Réunion avec le pôle informatique

L'environnement technique de la plateforme OpenEdition

Les informaticiens soulignent d'entrée de jeu le risque que représente une méconnaissance de l'environnement technique et de l'usage de la plateforme, ainsi que la difficulté à se former eu égard aux multiples usages de structuration des données en vigueur.

Les différents standards utilisés par OpenEdition sont OAI-PMH, Mets (Metadata Encoding & Transmission Standard, un format xml mis en place par la *Library of Congress* pour décrire des archives, l'arborescence, ainsi que les métadonnées descriptives sur DublinCore). Les portails Persée et Cairn utilisent Mets, ce qui rend possible l'interopérabilité des données.

La plateforme dispose de quatre types d'index :

- **IndexDocument** liste les formats des documents : chapitre de livre, article, billet, numéro, collection de numéros, événements scientifiques, etc. pour l'ensemble des contenus publiés ;
- **IndexPublisher** contient des informations sur l'ensemble des notices de revues au catalogue, ainsi que pour les carnets ; on trouve également une notice par éditeur d'OpenEditionBooks ;
- **IndexCalenda** est spécifique aux événements de Calenda qui sont aussi dans Indexdocument. Toutefois ils sont décrits de manière plus précise.
- Un dernier index est un **index d'auteurs** qui « récupère » tous les auteurs des quatre plateformes (à noter que, pour *Hypothèses*, ce n'est pas très fiable).

Le moteur de recherche d'OpenEdition est basé en interne sur une application nommée CORE, qui utilise le moteur d'indexation **solR**. Les commentaires n'y sont pas repris. Les blogs qui ne sont pas au catalogue ne sont pas non plus indexés dans ce moteur de recherche. Le panorama des interfaces produites (OAI PMH, Marc21...) peut être consulté sur la cartographie du rapport d'activité 2015 d'OpenEdition, p. 49.

Il existe une interface d'administration de solR qui permet de requêter dans les index, de manière très rapide. Il est en outre possible de dupliquer les données d'OpenEdition à un moment donné, et de

travailler ensuite sur ce snapshot. SolR fonctionne avec la [syntaxe de requête « lucene »](#), dont il faut connaître le schéma pour enquêter.

SolR permet de croiser les besoins : p. ex. de rechercher un CR, sur telle thématique, dans telle discipline, etc. Il intègre davantage de facettes de recherche, dont peu sont finalement accessibles par le moteur de recherche du portail (qui ne permet que l'interrogation de l'une ou l'autre facette).

Les thématiques attribuées aux documents sont héritées des classements d'OpenEdition, reposant sur les catégories de Calenda. Il existe trois classements généraux :

- **Index OpenEdition** provient de la récupération des catégories thématiques de Calenda (« thèmes ») ;
- **Index ISI** provient de la catégorisation par « disciplines » ;
- **Index Bisac** qui concerne plus spécifiquement les livres, est utilisé uniquement sur OpenEditionBooks ; il reprend des catégorisations standardisées, utilisées en librairie (sur le site d'Amazon par exemple).

Ces index sont également exploitables via solR et/ou les interfaces produites (OAI-PMH, Lonix pour les livres et librairies ; MARC pour les bibliothèques). Les classements sont effectués : sur les livres, par l'équipe d'OpenEdition, sur les revues et les carnets, par les utilisateurs qui les créent ; mais ils font ensuite l'objet d'une relecture et d'une validation par l'équipe.

Pour les besoins de sa recherche, I. MAYEUR souhaiterait accéder aux listes de diffusion du comité scientifique ayant servi, en interne, à la constitution des catégories disciplinaires en usage sur la plateforme. Elle s'enquiert de la possibilité de récupérer les listes sous un format exploitable, afin d'analyser qualitativement l'évolution des arguments ayant mené à l'évolution de l'objet « revue ». Il n'est pas certain que la liste de diffusion soit exportable ; par contre, B.-O. DOZO peut se charger de créer un accès pour I. MAYEUR à la liste privée, moyennant l'accord des autres membres de la liste et le respect du secret de la correspondance dans l'exploitation des données (anonymisation etc.).

I. MAYEUR voudrait également connaître les possibilités d'exportation des données du catalogue au format SQL (par exemple), de manière à pouvoir travailler sur une base de données collectant les informations relatives aux revues/carnets de recherche, textes de présentation, thématiques/disciplines, dates de création/du dernier post... Il apparaît toutefois qu'une telle exportation ne serait pas chose aisée ; il n'est par ailleurs pas possible de récupérer les classifications établies par les usagers des carnets de recherche. I. MAYEUR estime que cela fait partie des éléments qu'elle peut collecter par elle-même au départ du catalogue public.

Statistiques de consultation sur OpenEdition

Il existe plusieurs interfaces pour la consultation des statistiques d'utilisation :

- **Awstat** utilise les logs comme source de données (adresses IP + pays de consultation via GeolIP) ;
- **Graylog** fonctionne de manière analogue à l'index document, si ce n'est que les documents sont ici des logs de connexion : adresse IP : user agent, URL accédés... ;
- **usages.openedition.org** fournit des statistiques d'usage pour les contenus d'OpenEdition, à savoir les fréquentations de tous types et de tous formats par unité. Les données sont exportables en Excel ; mais n'intègrent pas la notion d'IP. On identifie par contre les bibliothèques qui s'abonnent aux contenus d'OpenEdition. Ces statistiques existent également pour *Hypothèses* : **usages.hypotheses.openedition**.

Il apparaît que les parcours d'utilisateur sont difficiles à reconstituer par ces biais. En outre, il est possible d'accéder à des informations génériques par rapport à une revue ou un blog, mais pas à un article. Le défi serait d'extraire des informations pertinentes à partir de statistiques assez pauvres.

I. MAYEUR pourrait, par un accès à l'interface d'administration de solR, accéder à l'ensemble des documents d'OpenEdition. Il lui serait également nécessaire d'obtenir une autorisation pour consulter les usages par graylog, qui n'est utilisé qu'en interne à des fins « système ». B.-O. DOZO précise que, dans un premier temps, il s'agira uniquement pour I. MAYEUR d'effectuer des repérages et non de produire une recherche. Techniquement, il serait réalisable de « photographier » les logs sur une période donnée et les exporter dans une base de données ; ce n'est toutefois pas le genre d'objets sur lesquels I. MAYEUR va travailler. Une possibilité serait de faire traiter ces informations par des professionnels de l'information-documentation, avec des étudiants orientés informatique, dans le cadre d'une recherche doctorale.

La solution retenue (moyennant les autorisations d'OpenEdition pour la consultation des données) est donc **l'installation d'un ordinateur à l'Université de Liège, fonctionnant avec l'environnement Linux. Les données resteront hébergées à Marseille, et la connexion sera établie via un tunnel SSH.** Par ailleurs, I. MAYEUR envisagera un prochain séjour à Marseille afin de se former à l'utilisation de solR. Une documentation sur le moteur de recherche est en train d'être développée.

Après le lunch, une **dernière réunion** a lieu de 13h30 à 15h30 avec M. DACOS, E. FAATH et P. BELLOT.

Débriefing et dégagement des perspectives de recherche

B.-O. DOZO résume les pistes dégagées par les réunions autour du projet de thèse d'I. MAYEUR, qui se concentre sur les **usages appréhendés par les objets textuels et documentaires** ; soit une approche davantage **qualitative** des contenus de la plateforme. Il pourrait être utile, peut-être en collaboration avec P. BELLOT, de trouver une personne disposant de compétences en matière de traitement des logs et en ingénierie documentaire, afin de compléter l'approche qualitative.

F. PROVENZANO pose la question de l'évolution diachronique de la plateforme et du développement de ses différents portails. B.-O. DOZO estime quant à lui qu'OpenEdition doit se concevoir avant toute chose comme une synchronie, sous peine de déboucher sur une analyse non pertinente. M. DACOS conseille cependant de porter attention au caractère évolutif de la plateforme dans le traitement des données : ainsi, on veillera soit à conserver les mêmes bornes chronologiques, soit on assumera le caractère mouvant des objets (comme c'est le cas pour les catégorisations). B.-O. DOZO ajoute qu'OpenEdition repose sur un état mixte, une forme de continuum agrégé. Dans ce contexte, le type de savoir à produire ne différera pas sensiblement de ce qui s'est produit dans les autres strates temporelles – ce qui permet d'ailleurs de limiter le champ de questionnement.

M. DACOS précise qu'une spécificité majeure de la plateforme est l'attention portée à son **appropriation** par les usagers, notamment par l'organisation de formations gratuites à leur intention. Dans le cadre de la thèse d'I. MAYEUR, les contenus de la plateforme constituent un corpus donné, mais tout à fait pertinent pour une étude des usages et mécanismes d'appropriation.

Ainsi, les archives des projets de financement peuvent constituer une source utile de contextualisation. Ces documents témoignent d'une évolution des discours de posture d'OpenEdition : les enjeux se sont déplacés, les interlocuteurs (pouvoirs de financement) ont changé. Il convient toutefois de rester attentif quant au fait que ces données ne sont en aucune manière publiques, et sont concernées par le **secret industriel**. L. GEORGE, chargée de communication, peut retrouver les traces de ces archives ;

elle a par ailleurs rédigé une sorte de storytelling d'OpenEdition qui pourra être communiquée [réf. à compléter].

B.-O. DOZO ajoute deux autres sources : CAVALLO 2009 et LEMERCIER 2009 M. DACOS trouverait par ailleurs utile qu'I. MAYEUR participe à un certain nombre de conseils scientifique d'OpenEdition, où sont arbitrées des décisions éditoriales quant au contenu de la plateforme. Il mentionne aussi le carnet **Leo**, carnet de recherche d'OpenEdition préalable à l'ouverture de celui du Lab (**Openlab**), où sont groupées les actualités que la plateforme désireait communiquer. Il renseigne également **Oep**, une version en anglais de Leo, avec un nombre plus réduit d'articles. Enfin, il est possible de trouver sur **InternetArchive** des traces de l'évolution diachronique de la plateforme : on y consultera par exemple le snapshot d'un texte de présentation en 2009. Il peut aussi être pertinent de s'intéresser aux évolutions terminologiques, par exemple l'abandon progressif du terme « portail » au profit de « plateforme ».

B.-O DOZO suggère à cet égard de recourir aux propositions théoriques de G. GENETTE (1979, 1982, 1987) sur le paratexte et le métatexte, afin de consolider l'ancrage philologique de la thèse et de lier l'étude des éléments contextuels à la circulation des savoirs au sein de la plateforme OpenEdition. F. PROVENZANO ajoute que l'appropriabilité des savoirs constitue bien une question centrale de la thèse, articulant la politique qu'OpenEdition entend encourager et la manière dont les revues hébergées constituent et animent leur champ disciplinaire. La thèse pourrait ainsi mettre en évidence le fait qu'une politique scientifique centrée sur la valeur de l'appropriabilité est susceptible d'entrer en résonance avec la manière dont une discipline fonctionne en interne (ce qui s'observe par le biais des commentaires, recensions, etc.). B.-O. Dozo ajoute que de ce point de vue, les revues *VertigO* et *Comicalités* constituent des exemples intéressants : la première, par le déplacement d'une discipline catégorisée en sciences vers les sciences humaines, la seconde, par l'établissement d'un champ disciplinaire peu légitime. Le dossier rentré par *Comicalités* pour candidater à la publication sur la plateforme est à cet égard très instructif. Ces dossiers peuvent représenter une autre trace paratextuelle à mobiliser ; toujours confidentielle par ailleurs. F. PROVENZANO pense qu'une étude est à faire sur les types de **valeurs** mises en avant dans ce dossier, afin de déterminer la manière dont elles entrent en résonances avec le projet éditorial d'OpenEdition.

M. DACOS avance d'autres pistes bibliographiques : l'étude réalisée par G. CHARTRON et M. MINON (2005), ainsi qu'un audit du CNRS de 2007 qui a partiellement fuité dans le *Figaro*. Il existe aussi une étude du JFII, groupe qui associe en France les industriels avec les acteurs publics [réf. à compléter].

B.-O. DOZO confirme donc qu'il existe bien un paratexte, qui méritera d'être pris en compte pour situer l'objet au moment de la justification par I. MAYEUR du choix de corpus. Ensuite, le deuxième volet de l'étude consistera à montrer comment OpenEdition contribue à **modifier les pratiques de circulation de savoir**. À cet égard, le passage au *Freemium* a changé la donne, étant donné qu'OpenEdition s'est allié les bibliothèques. La mise en place de la politique du *Freemium* fait suite à un rapport d'Emma Dexter, qui a interrogé les bibliothèques ; la totalité de ce rapport n'est pas public, mais les résultats ont été publiés dans des articles [réf. à compléter]. Il peut également se révéler utile de déterminer la manière dont OpenEdition s'est positionné dans le débat français entourant la loi numérique.

De là, il faudra questionner dans la thèse le pourquoi de cette plateforme, la représentativité de ce corpus donné dans ce paysage de l'édition scientifique électronique en sciences humaines (pagerank, nombre de document, présentation...). Il existe des outils dont dispose OpenEdition pouvant être mis à la disposition d'I. MAYEUR à cette fin, par exemple **alexa.com**, qui est un service payant mais pour lequel OpenEdition possède un abonnement. Alexa.com estime la fréquentation des sites dans le

monde et permet une veille concurrentielle ; il fournit ainsi des éléments permettant d'objectiver un positionnement sur le web, de donner des chiffres issus d'une interface tierce.

D'autres indicateurs utiles seraient les chiffres de fréquentation du portail par les bibliothèques ; toutefois celles-ci refusent de les donner. On risquerait là de glisser vers une étude des modèles économiques, ce qui est assez annexe au projet de thèse. B.-O. DOZO juge plus intéressant **d'analyser des discours institutionnels qui apparaissent dans les éléments paratextuels** que de faire de l'étude économique. F. PROVENZANO ajoute qu'il s'agit à présent de savoir dans quel ordre I. MAYEUR va procéder pour dégager son corpus, et comment vont s'organiser les étapes suivantes de la collaboration.

Un dernier élément important est la manière dont on labellisera l'apport d'OpenEdition dans les éventuelles publications à venir de la doctorante. La meilleure solution sera de discuter ensemble, au cas par cas, de la manière dont on procédera. En tout état de cause, la **collaboration se fait avec la structure OpenEditionLab**, et non OpenEdition.

Les **perspectives de recherche** ainsi dégagées pour la thèse d'I. Mayeur renouent davantage avec des **préoccupations philologiques**, éclairant les contenus textuels de la plateforme à partir des éléments de paratexte. L'analyse qualitative des données pourra se voir complétée par des approches quantitatives telles que l'analyse textométrique de plus grands ensembles textuels.

Récapitulatif des ressources disponibles

Données

- Base de données SOLR permettant l'accès aux contenus de *Revue.org* et *Hypothèses* ;
- Statistiques twitter établies par Élodie FAATH ;
- Statistiques d'utilisation de la plateforme : **usages.openedition.org** ;
- Statistiques relatives aux logs de connexion : **logs.openedition.org** ;
- Accès à **Graylog** et **Awstat** pour la visualisation de ces mêmes statistiques ;
- Accès à **alexa.com**.

Matériaux paratextuels

- Rapports d'activité d'OpenEdition ;
- Archives des demandes de financement et évaluations ;
- Liste de diffusion du conseil scientifique d'OpenEdition ;
- Carnets de recherche Leo, OpenEditionLab etc ;
- Participation d'OpenEdition aux débats entourant la loi numérique ;
- Dossiers de candidature des revues à *Revue.org*.

Personnes-ressources au sein d'OpenEdition

- BELLOT Patrice, LSIS-OpenEdition ;
- DACOS Marin, directeur du CLEO ;
- FAATH Élodie, chargée de la coordination des projets recherche, développement et innovation ;
- GEORGE Lisa, chargée de communication ;
- RIVIÈRE Jean-François, responsable du pôle informatique ;
- ORBAN Mathieu, membre du pôle informatique.

Bibliographie succincte du projet de thèse

AMOSSY, Ruth. 2010. *La présentation de soi: ethos et identité verbale*. Paris: Presses universitaires de France.

BERTHOU, Benoît. 2009. « La publication numérique en sciences humaines et sociales : vers un nouveau modèle de diffusion du savoir? » *Mémoires du livre* 1 (1). <https://www.erudit.org/revue/memoires/2009/v1/n1/038632ar.html>.

BONACCORSI, Julia, et Sarah LABELLE. 2010. « Autorité documentaire et pensée (inter)disciplinaire ». Colloque international «Edition et publication scientifique», mars 2010, Avignon.

BORDIER, Julien. 2016. « Évaluation ouverte par les pairs : de l'expérimentation à la modélisation ». <https://hal.archives-ouvertes.fr/hal-01283582/document>.

BRIATTE, François. 2015. « Hypothésosphère : le réseau des blogs d'Hypothèses, (presque) cinq ans plus tard ». *Polit'bistro : des politiques, du café*. <http://politbistro.hypotheses.org/2737>.

CAVALLO, Delphine. 2009. « Revues.org : l'invention de l'édition électronique scientifique, entre libre accès et modèle économique pérenne ». *Mémoires du livre* 1 (1). <https://www.erudit.org/revue/memoires/2009/v1/n1/038638ar.html>.

COMPAGNON, Antoine. 1979. *La Seconde main ou le Travail de la citation*. Paris: Seuil.

CORDONNIER, Sarah. 2015. *Trajectoire et Témoignage: Pour une réflexion pluridisciplinaire*. Paris: Editions des archives contemporaines.

DACOS, Marin. 2016. « La mécanique des fluides ». Billet. Consulté le février 10. <http://bn.hypotheses.org/162>.

DACOS, Marin, et Pierre MOUNIER. 2010. « Les carnets de recherche en ligne, espace d'une conversation scientifique décentrée ». In *Lieux de savoir. 2. Gestes et supports du travail savant*, édité par Christian Jacob, 2:N/A. Paris: Albin Michel. http://archivesic.ccsd.cnrs.fr/sic_00439849/document.

DAVALLON, Jean, Marie DESPRÉS-LONNET, Yves JEANNERET, Joëlle LE MAREC, et Emmanuël SOUCHIER. 2003. *Lire, écrire, récrire : Objets, signes et pratiques des médias informatisés*. Études et recherche. Paris: Éditions de la Bibliothèque publique d'information. <http://books.openedition.org/bibpompidou/394>.

DEBAZ, Josquin. 2010. « Essai de cartographie des carnets de recherche ». Billet. *Hypothésosphère*. décembre. <http://socioargu.hypotheses.org/1921>.

FAATH, Élodie. s. d. « Cartographie de la circulation de l'information scientifique sur un espace éditorial défini : l'exemple d'OpenEdition Sphère ».

FAATH, Élodie, Chahinez BENKOUSAS, Hussam HAMDAN, Patrice Bellot, Marin DACOS, et Frédéric BÉCHET. 2013. « Machine Learning and Text Mining for Retrieving Reviews of Books ». Billet. *OpenEditionLab*. <http://lab.hypotheses.org/867>.

GENETTE, Gérard. 1979. *Introduction à l'architexte*. Paris : Seuil.

Id., 1982. *Palimpsestes*. Paris : Seuil.

Id., 1987. *Seuils*. Paris : Seuil.

GRANJON, Fabien, et Christophe MAGIS. 2016. « Critique et humanités numériques ». *Variations. Revue internationale de théorie critique*, n° 19(avril). <https://variations.revues.org/748>.

JEANNERET, Yves. 2014. *Critique de la trivialité: Les médiations de la communication, enjeu de pouvoir*. Paris: Editions Non Standard.

JEANNERET, Yves, et Emmanuël Souchier. 2005. « L'énonciation éditoriale dans les écrits d'écran ». *Communication et langages* 145 (1): 3-15. doi:10.3406/colan.2005.3351.

LEMERCIER, Claire. 2009. « Classer les revues de sciences humaines et sociales : comment ? pour quoi ? » In *La valeur de la science. Pourquoi évaluer la recherche ?* Liège, Belgium. <https://halshs.archives-ouvertes.fr/halshs-00521517>.

MAINGUENEAU, Dominique. 2013. « Genres de discours et web : existe-t-il des genres web ? » In *Manuel d'analyse du web en Sciences Humaines et Sociales*, 74-93. Paris: Armand Colin.

MCKENZIE, D.-F., et Roger CHARTIER. 1991. *La Bibliographie et la sociologie des textes*. Cercle de la Librairie.

MERCKLÉ, Pierre. 2011. « L'Hyposphère ». [pierremerckle.fr. http://pierremerckle.fr/2011/06/lhyposphere/](http://pierremerckle.fr/2011/06/lhyposphere/). Consulté le 21 avril 2016.

MINON, Marc, et Ghislaine CHARTRON. 2005. « Etat des lieux comparatif de l'offre de revues SHS, France-Espagne-Italie. Etude réalisée pour le Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche ». Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche MESR. http://archivesic.ccsd.cnrs.fr/sic_00001561/fr/.

PAVEAU, Marie-Anne. 2015. « Ce qui s'écrit dans les univers numériques ». *Itinéraires. Littérature, textes, cultures*, n° 2014-1(février). doi:10.4000/itineraires.2313.

PUSCHMANN, Cornelius, et Marco BASTOS. 2015. « How Digital Are the Digital Humanities? An Analysis of Two Scholarly Blogging Platforms ». *PLOS ONE* 10 (2): e0115035. doi:10.1371/journal.pone.0115035.

SAEMMER, Alexandra. 2015. *Rhétorique du texte numérique: figures de la lecture, anticipations de pratiques*. Lyon: Presses de l'Enssib.

SOUCHIER, Emmanuel. 1998. « L'image du texte : pour une théorie de l'énonciation éditoriale ». *Les cahiers de médiologie* N° 6 (2): 137-45.

VITALI ROSATI, Marcello. 2016. « What is editorialization? » *Sens public*, janvier. <http://www.sens-public.org/article1059.html>.

Actions à mener suite à la mission

- Obtenir les **autorisations du Conseil Scientifique** d'OpenEdition afin de permettre à I. MAYEUR d'accéder aux contenus de sa liste de diffusion ;
- Obtenir les **autorisations du CLEO** pour l'accès à une base de données SOLR permettant d'interroger les contenus de la plateforme OpenEdition à une date déterminée ;
- Installer à Liège d'un **portable tournant sur Linux**, destiné à accueillir la connexion par tunnel SSH ;
- Créer un **carnet de thèse** sur *Hypotheses*, destiné à accompagner le projet de thèse d'I. MAYEUR et qui permettrait de conserver des traces de son avancement ;
- Réfléchir à une éventuelle **charte de collaboration**, destinée à développer de bonnes pratiques, de manière à ce que les recherches issues de données modélisées par OpenEdition soient aisément identifiables (ex. : citation en tant que co-auteur si une publication exploite des jeux de données provenant d'OpenEdition) ;
- Définir plus précisément le **corpus de la thèse** : cibler les contenus textuels et leur catégorisation sur la plateforme, puis définir des sous-corpus d'investigation correspondant aux questions posées ; à interroger ensuite avec les outils de recherche et de visualisation des sources.