

Compte-rendu de l'assemblée générale ordinaire de l'association C.A.P.S.H.

Antoine Amarilli

20 septembre 2016

L'assemblée générale de l'association C.A.P.S.H. s'est tenue au cours de la soirée du lundi 20 septembre 2016 au domicile d'Antoine Amarilli à Montrouge (92), à partir de 19h. Sont présents :

- Antoine Amarilli
- Marie Farge
- Antonin Delpeuch
- Marc Jeanmougin
- Pablo Rauzy
- Ryan Lahfa
- Patricia Mirabile

Antonin Delpeuch endosse la responsabilité de président de séance et Antoine Amarilli celle de secrétaire de séance. Les points suivants figurent à l'ordre du jour :

- Bilan moral
- Bilan financier
- Partenariat CCSD/Couperin
- Résultats du financement PSL
- Relations avec Academia.edu
- Relations avec ORCID
- Gestion des articles scientifiques manquants
- Organisation centralisée ou fédérée
- OAbot
- OA Week
- Démonstration du dépôt dans HAL
- Open Science Conference
- Listes de diffusion
- Point sur la loi Lemaire
- Dissemmin et données ouvertes
- Plateformes de discussion
- Organisation de réunions régulières

Bilan moral. Le bilan moral est présenté par Antonin Delpeuch. Il fait brièvement état des principales actions entreprises par l'association au cours de l'année écoulée, et de leurs retombées. Le bilan est tacitement approuvé par les membres présents.

Bilan financier. Le bilan financier est présenté par Antoine Amarilli. Il s'appuie sur des documents préalablement portés à la connaissance des membres de l'AG par voie électronique, et dont la teneur suit.

Au 12 septembre, le solde du compte bancaire de l'association se porte à 183.99 EUR. Ses recettes mensuelles (dons récurrents) se portent à 15 EUR, et ses dépenses mensuelles à 86.36 EUR.

Sur la période courant du 1^{er} septembre 2015 au 31 août 2016, les recettes totales se montent à 877.18 EUR, dont 155 EUR de dons récurrents (deux donateurs), et 722.18 EUR de dons ponctuels reçus par virement bancaire : le 2015-10-05 (50 EUR), 2015-10-22 (300 EUR), 2015-10-29 (67.75 EUR), 2015-11-02 (27.18 EUR), 2015-11-10 (30 EUR), 2015-11-25 (97.25 EUR), 2016-02-18 (150 EUR).

Les dépenses totales se montent à 611.83 EUR, dont une dépense ponctuelle de 32.40 EUR pour renouvellement de nom de domaine auprès de Gandi, et des dépenses totales de 579.43 EUR de location de serveurs auprès de Online.

Partenariat CCSD/Couperin. Interrogé sur la santé financière de l'association, Antonin Delpeuch rappelle que le partenariat avec le CCSD, déjà signé, prévoit un financement de 9 000 EUR pour l'association, dont le versement devrait intervenir prochainement. Il est calculé que ce financement pourra couvrir les frais de fonctionnement actuels de l'association, sur la base des dons récurrents actuels, pour une période de plus de dix ans.

Pour réduire les dépenses, Antonin Delpeuch mentionne également la possibilité de renoncer à la location d'un des deux petits serveurs actuellement loués par l'association, une fois que la migration des services vers le nouveau serveur (de plus forte capacité) aura été accomplie.

Résultats du financement PSL. L'AG regrette de ne pas avoir encore eu connaissance des résultats de la candidature de l'association à l'appel à projet PSL Explore, et reporte la discussion de ce point à une date ultérieure.

Relations avec Academia.edu. L'AG salue l'ouverture par Academia.edu d'une interface OAI-PMH permettant le moissonage d'articles. Antonin Delpeuch rappelle que Academia.edu a également choisi de permettre aux utilisateurs atteignant une de leurs pages depuis Wikipédia d'accéder au texte intégral des articles hébergés sans avoir besoin de se créer un compte.

Antonin Delpeuch mentionne la situation avec ResearchGate, dont le crawl n'est pas fonctionnel à l'heure actuelle.

Relations avec ORCID. Marie Farge mentionne les discussions qu'elle a menées avec ORCID. Certaines réserves sont émises au sujet de la collecte d'informations, qu'ils entreprennent suivant différents fronts (par exemple, source de financements), et sur leur statut de *not-for-profit* qui ne semble toutefois pas les mettre entièrement à l'abri d'une appropriation commerciale. Sont toutefois mentionnées les garanties de gouvernance rapportées à Marie Farge. L'importance d'ORCID et l'existence de solutions alternatives sont discutées, mais Marie Farge rappelle qu'ORCID est la solution commune vers laquelle ont convergé les acteurs existants, et qu'il apparaît donc difficile de la remplacer.

En ce qui concerne l'utilisation d'ORCID par Dissemin, Antonin Delpeuch rappelle qu'ORCID est utilisé pour externaliser le problème de l'authentification des chercheurs, et celui de la désambiguation d'articles : suivant le plan actuel, les utilisateurs actuels qui ne sont pas associés à un compte ORCID pourront simplement énumérer leurs articles par le biais d'une recherche textuelle, avec de possibles problèmes d'homonymie. Une fois associés à un compte ORCID, le plan prévoit de permettre aux utilisateurs de modifier leur liste de publications sans les renvoyer vers l'interface ORCID tierce pour ce faire. En revanche, ceci ne semble pas possible sans devenir membre d'ORCID. Après étude des tarifs proposés sur le site d'ORCID, l'AG estime qu'une adhésion à ORCID n'est pas financièrement viable aujourd'hui. Antonin Delpeuch fait état de l'échec de tentatives antérieures d'obtenir malgré tout un accès préférentiel aux fonctionnalités nécessaires.

L'AG discute la possibilité d'une adhésion à ORCID par Couperin, qui pourrait alors faire bénéficier Dissemin de leur accès. Malheureusement, l'AG estime que la structure complexe de Couperin, et les montants et délais en jeu, font que cette perspective ne pourra probablement pas régler le problème à court terme.

Antoine Amarilli fait remarquer le déséquilibre apparent de la situation actuelle, où l'association devrait payer ORCID pour avoir le droit de leur fournir des données. L'AG demande si d'autres acteurs se trouvent dans une situation analogue à la nôtre. Antonin Delpeuch rappelle que les outils de gestion de recherche vendus aux universités les placent dans une situation similaire à la nôtre ; à leur échelle, cependant, le montant de la cotisation à ORCID ne représente pas un problème de la même ampleur.

L'AG propose ainsi une solution où les éditions de publications par les utilisateurs seraient conservées localement par Dissemin, mais ne seraient pas propagées à ORCID, faute d'une interface pour ce faire. L'AG mentionne également la possibilité de permettre aux utilisateurs une intégration manuelle de leurs données dans ORCID, par exemple par export BibTeX ou par l'envoi de courriels. D'autres points techniques sont discutés, notamment l'impact d'une édition effectuée par un utilisateur sur la présentation du papier concerné dans la liste de publications d'autres utilisateurs.

Gestion des articles scientifiques manquants. Ryan Lahfa sonde l'AG quant à la réponse opportune à apporter aux gestionnaires de dépôts qui demandent à Dissemin de les indexer. Antonin Delpeuch rappelle que ces dépôts doivent être inscrits dans BASE.

Marie Farge soulève la question de l'identification des sources en *green open access* et en *gold open access*. Elle souligne les limites de l'approche actuelle, et s'interroge sur la pertinence de conserver ces indicateurs, au vu des erreurs commises et de la possibilité

de mésinterprétation de ces statistiques. Antonin Delpeuch répond que cette question est actuellement en discussion, suite à une remontée utilisateur.

Il est proposé par l'AG de permettre aux utilisateurs de Dissemin de renseigner les détails d'articles inconnus des sources bibliographiques, pour les inclure dans une liste de publications. Un risque serait toutefois que des utilisateurs usurpent les publications de tiers. Marie Farge fait état de la frustration récurrente des chercheurs, lorsqu'ils constatent que des informations grossièrement erronées au sujet de leurs articles sont ajoutées à des référentiels bibliographiques, le plus souvent sans contrôle et sans possibilité pour le chercheur d'effectuer des corrections.

Antoine Amarilli soulève la question de l'indexation Web, pour extraire les articles de recherche qui ne sont pas présents dans les sources bibliographiques de Dissemin. Antonin Delpeuch distingue trois utilisations possibles de l'indexation : premièrement, la validation des informations de disponibilité du texte intégral mentionnées par nos sources bibliographiques, à effectuer par un téléchargement manuel du fichier PDF ; deuxièmement, un usage restreint de l'indexation, sur des dépôts manuellement identifiés et dépourvus d'interface OAI-PMH, pour en extraire les articles ; troisièmement, une utilisation large de l'indexation, similaire à celle de Google Scholar, pour identifier davantage d'articles. Il souligne néanmoins les difficultés techniques que rencontrerait une pareille indexation.

Marie Farge rappelle toutefois l'enjeu du problème des papiers manquants. Elle indique qu'il pourrait s'avérer délicat de présenter Dissemin à des utilisateurs potentiels, en raison du risque d'échec si le système ne répertorie aucune publication du chercheur concerné. Elle indique que, dans certains pays, ce problème se pose de manière récurrente.

Organisation centralisée ou fédérée. Ryan Lahfa attire l'attention de l'AG sur le changement progressif de stratégie de Dissemin, et sur ses conséquences techniques. La vision originale de Dissemin était de proposer un système fédéré, qui pouvait être installé indépendamment par chacun des institutions concernées. Néanmoins, étant donné la nécessité d'accomplir des tâches de moissonnage gourmandes en ressources, Dissemin s'oriente davantage vers un modèle centralisé, où une instance centrale (gérée par l'association) se chargerait seule de l'exécution de ces tâches.

Après discussions, un modèle hybride est proposé, où l'instance centrale se chargerait des tâches lourdes ; étant bien entendu que, le code logiciel étant disponible sous licence libre, ces tâches pourraient cependant être exécutées indépendamment par tout utilisateur intéressé. L'instance centrale pourrait alors proposer une interface de programmation (API) permettant à des instances "légères" d'accéder au résultat des opérations. Ces instances permettraient aux institutions intéressées de pouvoir facilement installer Dissemin localement sur leurs machines, et leur permettrait également de personnaliser facilement l'interface, de l'intégrer à des portails d'authentification locaux, et d'en modifier le design conformément à leur image de marque. La question de la politique d'accès à l'API est discutée : celle-ci pourrait être payante, à titre de participation aux frais. Antoine Amarilli mentionne la possibilité de mise à disposition par les instances légères d'une partie de leurs ressources, en contrepartie d'un accès à l'API, afin d'être mises à

contribution pour les tâches lourdes.

L'AG évoque également la position à tenir pour répondre aux demandes d'universités ou de documentalistes qui désirent mettre en place une installation de Dissemin. À l'heure actuelle, il n'est pas décidé d'encourager cette pratique, étant donné le caractère préliminaire du prototype actuel d'implémentation. Marie Farge rappelle toutefois que Dissemin répond à une véritable demande de la part de nombreux acteurs.

OAbot. Antonin Delpuch rappelle que l'objet d'OAbot est de travailler à l'édition automatisés des liens vers des articles scientifiques sur l'encyclopédie collaborative en ligne Wikipédia, afin de privilégier les versions disponibles en libre accès. Antonin fait brièvement état des réunions entreprises à ce sujet avec les wikipédiens intéressés, et avec CUNY, notamment par rapport à la proposition d'intégrer à OAI-PMH des informations plus riches sur la licence des articles (et non seulement sur leur disponibilité). L'enjeu de visibilité sur Wikipédia est notamment de nature à encourager les universités à œuvrer pour ajouter des liens vers les articles de leurs chercheurs sur Wikipédia.

Antonin rappelle l'état actuel de la discussion, qui porte notamment sur le graphisme de l'icône indiquant les copies en libre accès dans les références de Wikipédia. Il mentionne les difficultés techniques liées à l'édition de patrons (*templates*) utilisés sur de nombreuses pages Wikipédia, qui nécessite de faire appel à un administrateur.

OA Week. L'AG s'interroge sur la possibilité d'organiser des événements dans le cadre de l'OA Week. Les membres expriment des réserves quant à leurs disponibilités à cette fin, et quant à la proximité de l'échéance ; la possibilité que Dissemin s'intègre au programme d'autres institutions est notamment mise en avant. L'AG discute toutefois des possibilités d'organisation d'un hackathon autour du libre accès, par exemple en lien avec la mission Etalab ; mais cette proposition dépend entre autres des financements obtenus par l'association entre temps.

Démonstration du dépôt dans HAL. Antonin Delpuch fait une démonstration à l'AG du prototype actuel d'interface pour le dépôt dans HAL. Le champ "Abstract" est automatiquement rempli si possible, et la classification de l'article est inférée à partir de celle-ci si nécessaire. Antoine Amarilli interroge Antonin sur le champ permettant de déposer pour le compte d'un tiers, jugé inutile dans le cas d'utilisateurs identifiés par un compte HAL et déposant pour leur propre compte. Marie Farge rappelle les problèmes liés aux dépôts massifs effectués pour le compte d'un tiers.

Le champ "Affiliations", concernant uniquement le déposant, est autocomplété à partir de ce que l'utilisateur saisit. Marie Farge insiste sur la nécessité de permettre aux utilisateurs d'ajouter leurs propres institutions, par exemple par saisie de texte libre.

Pablo Rauzy soulève la question de la modération effectuée par HAL, qui implique que le dépôt est susceptible d'échouer après un certain délai. En conséquence, il apparaît nécessaire de proposer un retour aux utilisateurs sur le résultat de leurs dépôts dans HAL. Antoine Amarilli pose également la question de modifications ultérieures que le déposant, ou ses coauteurs, pourraient vouloir effectuer directement dans HAL. Antonin

Delpuech et Ryan Lahfa rappellent que HAL indique à Dissemin le résultat de l'action par courriel, et transmet également un mot de passe permettant de prendre contrôle de l'article sur HAL.

Deux solutions, possiblement complémentaires l'une de l'autre, sont étudiées par l'AG. En premier lieu, Dissemin peut faire suivre aux utilisateurs par courriel le résultat de l'action et le mot de passe communiqué par HAL. En second lieu, ces informations (et, dans l'intervalle, le nombre d'articles en cours de modération dans HAL) peuvent être communiquées à l'utilisateur par un bandeau sur le site. Il est mentionné que les mots de passe envoyés par HAL constitueraient des données sensibles et devraient être traitées en conséquence.

Open Science Conference. Marie Farge indique que Raghavendra Gadagkar donnera une conférence à l'ENS le 12 décembre, et que des manifestations liés à Dissemin pourraient s'intégrer à cet événement. Notamment, si Dissemin obtient le financement PSL Explore, de telles manifestations pourraient correspondre aux volets des animations liées au libre accès, que C.A.P.S.H. s'engagerait alors à organiser à l'ENS.

Listes de diffusion. Antonin Delpuech propose un récapitulatif des abonnés à la liste de diffusion interne de l'équipe, et procède à un rappel des bonnes pratiques de communication électronique liées à leur usage. Pablo Rauzy mentionne la possibilité d'apposer un préfixe à l'objet des messages transmis par la liste, afin d'éviter toute confusion.

Point sur la loi Lemaire. L'AG rappelle les enjeux de la loi Lemaire liés au libre accès et aux objectifs de l'association, et s'informe de son progrès actuel dans le processus législatif.

Dissemin et données ouvertes. Ryan Lahfa interroge l'AG sur les données gouvernementales dont C.A.P.S.H. pourrait avoir besoin, en vue de demander leur ouverture. Marie Farge mentionne les fiches bibliographiques soumises par les personnels CNRS dans le cadre de leurs activités. L'AG discute de la possibilité d'en obtenir communication par le biais d'un recours CADA, mais émet des réserves liées au caractère nominatif de ces documents.

Plateformes de discussion. Ryan Lahfa interroge l'AG quant à la possibilité d'utiliser de nouvelles plateformes de messagerie instantanée en ligne pour Dissemin, qui seraient plus faciles d'accès que la plateforme IRC actuellement utilisée. La question de l'accès à IRC est posée par certains membres de l'AG. Plusieurs solutions techniques sont proposées. La nécessité de diviser le canal de discussion IRC actuel en plusieurs canaux thématiques est également discutée.

Organisation de réunions régulières. L'AG discute de l'opportunité d'organiser de réunions périodiques, sujettes aux disponibilités de ses membres et à leurs contraintes

géographiques. L'AG se décide pour une réunion mensuelle, le premier jeudi du mois à 17h30.

L'ordre du jour étant épuisé, la séance est levée à 21h50.