

Compte-rendu du CERN Workshop on Innovations in Scholarly Communication (OAI9), 17-19 juin 2015, Genève

Anne Gentil-Beccot
Anne.Gentil-Beccot@cern.ch
Bibliothèque du CERN

Salomé Rohr
salome.rohr@cern.ch
Bibliothèque du CERN

La 9^e édition de la conférence OAI a eu lieu du 17 au 19 juin 2015, à Genève, dans les locaux de l'Institut des Hautes Etudes Internationales et du Développement et du Campus Biotech de l'Université de Genève. Pour cette édition, l'Open Access était évidemment présent, mais l'accent a surtout été mis sur ce qu'il y a au-delà : Linked data, Open Data, Open Peer Review, Open Science... En effet, l'Open Access n'est pas une fin en soi ; ce sont les services que l'on construit autour qui permettent de faire évoluer la communication scientifique. Par ailleurs, il est à noter que le thème de l'institution en tant qu'éditeur scientifique s'est invité au programme pour la première fois.

1. Présentation d'ouverture : Au-delà de l'Open Access, par Michael Nielsen

Michael Nielsen pose d'emblée la question « L'Open Access à quoi ? ». Pour lui, l'Open Access n'est pas seulement le libre accès aux articles de journaux scientifiques ; l'enjeu se situe au-delà. « L'Open Access a le potentiel d'amplifier l'intelligence individuelle et collective », mais seulement, soutient-il, si nous développons de nouvelles formes de médias. Il cite par exemple le notebook *ipython*([1](#)), qui permet d'agrèger, de visualiser des données et du code, mais également de partager ses découvertes. Un tel outil permet d'augmenter le nombre de personnes qui interagissent ensemble ; le savoir devient alors d'autant plus extensible. Par conséquent, l'Open Access n'est qu'une partie d'un tout. Pour lui, les politiques d'Open Access doivent être développées de sorte que l'innovation et l'expérimentation ne soient pas inhibées.

2. Session technique

La première session des conférences OAI est toujours centrée sur les aspects techniques. Cette année, l'accent a été mis sur la publication des données.

2.1. Un accès durable et permettant l'interrogation aux *Linked data*([2](#)), par Ruben Verborgh

Pour Ruben Verborgh, l'essentiel pour les *Linked data*, est qu'elles soient facilement intégrées et que l'on puisse les interroger via le web. De nombreuses données sont désormais rendues disponibles, mais la grande difficulté reste leur accessibilité. Il présente donc une nouvelle façon de publier ces données de façon à les rendre accessibles et interrogeables, tout en limitant les coûts inhérents à ces processus. Avec la technologie *Triple Pattern Fragments*([3](#)), les Serveurs offrent une interface qui permet aux Clients d'exécuter des commandes complexes efficacement. Cette technologie permet des recherches fédérées avec tous types de données. Le logiciel est disponible en Open Source sur *Github*([4](#)) et donc réutilisable par toute institution souhaitant mettre à disposition ses données.\$

2.2. Un réseau décentralisé pour publier des *Linked data*, par Tobias Kuhn

Tobias Kuhn souligne lui-aussi le fait que la publication scientifique aujourd'hui n'est plus centrée seulement sur l'objet 'article', mais doit intégrer l'ensemble des données qui sont le résultat des recherches scientifiques. Il présente quelques outils qui selon lui sont essentiels pour publier des données de façon efficace. Les nanopublications([5](#)) (plus petites unités d'information publiables) seront désormais au centre de la communication scientifique. Ces unités seront ainsi facilement citables et réutilisables. Il présente aussi le concept des *Trusty*

URIs(6) qui permet de citer un set de données de façon à assurer que ces données sont à tout moment vérifiables, non modifiées et pérennes. Pour lui, les nanopublications doivent être publiées de façon décentralisée(7). Enfin, il évoque le concept des *Science Bots*(8) qui pourraient participer au contrôle de qualité de ces processus.

2.3. “Reference rot” et annotations de liens, par Martin Klein

Martin Klein présente le projet Hiberlink(9) qui vise à quantifier le nombre de ressources citées dans des articles scientifiques mais qui ne sont plus disponibles, ce qu’il nomme *Reference rots* (erreurs 404, pages dont le contenu a changé, etc.). L’équipe a étudié un corpus d’articles publiés entre 1997 et 2012 sur arXiv, Elsevier et PMC ; à partir des données de cette étude, ils ont estimé que sur l’ensemble des publications en STM(10), 1 article sur 5 comporterait des références dont les liens sont brisés. Pour pallier à ce problème, il propose l’idée des *Robust links*(11) (liens consolidés). Le *Robust Link* implique plusieurs éléments : la page citée doit être archivée sur une archive web telle que *archive.is*(12) et le lien doit être tagué avec la date et l’heure de la version de la page citée ainsi que l’URI de la version archivée de la page.

2.4. Barrières et impacts

Dans cette session, les conférenciers nous ont présenté des outils et conseils pour tenter de briser les barrières qui s’opposent à l’Open Access et ainsi augmenter l’impact de la recherche scientifique.

2.5. L’Open Access Button, par Joseph McArthur

L’initiative “Open Access Button”(13) permet aux personnes confrontées à un *paywall*(14) pour accéder à un article scientifique de tenter de faire évoluer l’accès à la recherche en rapportant leurs histoires qui seront ensuite utilisées pour tenter de faire changer les systèmes de publication vers l’Open Access. Une fois l’application installée, le bouton permet aux chercheurs de signaler les *paywalls* et de demander l’accès au document en expliquant pourquoi il a besoin d’y accéder. Le bouton va alors en rechercher une version accessible. S’il n’en existe pas, un e-mail sera envoyé à l’auteur et la référence sera enregistrée dans la *wishlist* du chercheur.

2.6. Les jeunes chercheurs et l’Open Access, par Erin McKiernan

Confrontée aux problèmes d’accès à la littérature scientifique en Amérique latine, Erin McKiernan plaide en faveur de la publication en Open Access. Selon elle, les jeunes chercheurs sont préoccupés par :

- Les coûts de publication élevés : alors que certains journaux OA (15) ne font pas payer de frais ou ont des frais d’adhésion forfaitaires peu élevés ; par ailleurs, certaines institutions ou financeurs ont des fonds dédiés à la publication OA ; enfin, l’auto-archivage ne coûte rien.
- La qualité du Peer-Review : pourtant, il n’y a pas d’étude qui prouve que le Peer-Review est de moins bonne qualité dans les revues OA ; la preuve de la qualité réside dans les « Open Peer Reviews » où le chercheur peut lui-même se faire une opinion sur la rigueur du processus.
- Leur carrière : les perspectives d’emploi et les évaluations sont basées sur le facteur d’impact. Il est cependant possible de choisir des revues OA à impact modéré à élevé. L’outil *Journal Selector* de Cofactor(16) permet de choisir une revue en fonction aussi des critères d’accès. L’initiative DORA(17) plaide en faveur d’un changement des critères d’évaluation des chercheurs.
- L’opinion négative des co-auteurs et mentors : une explication des bénéfices et des différentes options de publication OA pourraient les faire changer d’avis.

2.7. Wikimedia et l'intégration de l'Open Access, par Daniel Mietchen

Wikimedia est étroitement liée à l'Open Access en ayant pour but de rendre la recherche librement accessible et réutilisable. Ses plateformes et outils pourraient amener les chercheurs à reconsidérer la publication de leurs articles vers un modèle ouvert. Les outils Wikimedia permettent d'intégrer du matériel Open Access et différentes interactions sont possibles, par exemple, la publication d'articles, la signalisation des sources Open Access sur Wikipédia, des liens vers des pages web archivées, les plateformes Wikimedia Commons ou Wikidata. La possibilité, notamment, de la visualisation interactive et des graphiques éditables, permet d'intégrer des éléments multimédia complexes.

3. Les processus en Open Science

Cette session s'est concentrée sur l'Open Science qui, au-delà de la publication en Open Access, s'intéresse à l'ouverture du processus de recherche dans sa globalité.

3.1. CHORUS : améliorer l'accès à la recherche, par Howard Ratner

L'organisation CHORUS([18](#)) a pour but d'améliorer la communication scientifique en facilitant, normalisant et résolvant les problèmes liés aux 5 étapes du processus de recherche : identification, découverte, conformité, préservation, accès ; et ce pour tous les acteurs du processus. CHORUS se base sur des collaborations avec des outils existants, par exemple, *Cross Ref's Fund REF*, *ORCID*, *CLOKKSS* ou *PORTICO*. Lors du processus de soumission, l'auteur saisit les informations d'identification sur la source de financement de sa recherche. Cela permet de tagguer l'article avec les identifiants *Cross Ref's Fund REF* et l'information est envoyée avec le DOI et les informations bibliographiques à *Cross Ref*. Les métadonnées bibliographiques des membres sont collectées par CHORUS et un tableau de bord([19](#)), une API([20](#)) et un moteur de recherche permettent leur recherche et exploitation.

3.2. Ouvrir les infrastructures et structures de recherche dans l'enseignement supérieur, par Tyler Walters

L'initiative SHARE([21](#)) a pour but de créer un set de données ouvertes sur les activités de recherche tout au long de leur cycle de vie ; cela a pour but d'accélérer et d'améliorer la recherche dans l'enseignement supérieur et de permettre l'innovation. Les métadonnées sur les publications et la documentation des activités de recherche de différentes sources sont collectées à travers SHARE, normalisées et ensuite notifiées grâce à un flux RSS aux utilisateurs du service. Pour contribuer, les organisations doivent s'enregistrer pour que leurs données soient moissonnées. Des APIs permettent le partage de matériel et la recherche des données. Les challenges à relever restent l'harmonisation et la qualité des métadonnées ainsi que les questions de droits sur le partage des métadonnées et la promotion de ce partage.

3.3. Changer les processus de recherche et passer de l'Open Access à l'Open Science, par Bianca Kramer et Jeroen Bosman

Changer les processus de recherche ne concerne pas les outils, qui montrent seulement comment les chercheurs travaillent, mais les chercheurs eux-mêmes. La science s'est transformée d'un mode cyclique simple à un processus complexe multi-cyclique, multi-ordonné et multi-itératif. La transition du schéma traditionnel vers l'Open Science devrait amener à des pratiques plus honnêtes/reproductibles, efficaces et ouvertes([22](#)). Une étude([23](#)) est actuellement menée sur les outils utilisés par les chercheurs sur 17 activités du processus afin de voir comment la communication scientifique a changé.

4. Assurance de la qualité

Cette session était entièrement dédiée au processus de Peer-Review, qui comme tous les aspects de la communication scientifique se doit d'évoluer avec les nouveaux besoins et enjeux.

4.1. Gérer le processus d'évaluation par les pairs (peer review) à grande échelle, par Damian Pattinson

PLOS One⁽²⁴⁾ est une publication en libre accès multidisciplinaire ; en 2014, 32,000 articles ont été publiés. Avec un tel corpus, PLOS One est l'une des plus grandes publications scientifiques actuelles. Pour le processus d'évaluation des articles par les pairs, la revue combine le travail du bureau éditorial interne et celui, volontaire, d'éditeurs académiques et de chercheurs. Damien Pattinson fait remarquer que la recherche est de plus en plus interdisciplinaire, ce qui n'est pas forcément le cas des personnes qui évaluent les articles. Pour lui, le processus d'évaluation doit donc s'ouvrir aux lecteurs pour être plus efficace et s'adapter aux nouveaux enjeux de la recherche scientifique.

4.2. Peerage of science : une initiative issue de la communauté scientifique pour améliorer le processus d'évaluation par les pairs, par Janne-Tuomas Seppänen

Peerage of Science⁽²⁵⁾ est une initiative qui permet de rendre le processus de peer-review indépendant de la soumission à une revue. Un auteur peut en effet soumettre un article sur le site ; l'article sera révisé par un pair (dont le profil a été validé) ; et la révision elle-même sera évaluée. Le résultat est rendu disponible à toutes les revues scientifiques adhérentes à l'initiative. Un nouveau projet est en cours : *my.peerageofscience.org* où il est possible de visualiser les profils des pairs ainsi que leur performance et leur popularité.

4.3. Publons : donner du crédit au processus d'évaluation par les pairs, par Andrew Preston

Une initiative similaire à la précédente est *Publons.com*⁽²⁶⁾. Cette plateforme offre aux éditeurs les outils pour identifier et contacter les personnes qui pourront faire la révision des articles. Ces personnes trouvent un bénéfice à effectuer ce travail, car leurs évaluations sont référencées et évaluées.

5. L'institution comme maison d'édition

Il s'agit là d'un nouveau sujet pour la conférence OAI. Sa présence ici révèle l'importance du phénomène dans le monde académique.

5.1. Mettre le monde académique au cœur de l'édition scientifique, par Catriona Maccallum

La recherche n'est désormais plus un cycle mais un réseau dans lequel de très nombreux acteurs sont impliqués. Bien sûr, il y a l'enjeu du libre accès au contenu même des articles, mais il s'agit d'aller au-delà et d'interconnecter tous ces acteurs pour développer des services autour du contenu. En cela, les institutions ont un grand rôle à jouer, en particulier en développant la transparence à tous les niveaux (Peer-Review, coûts de publication, etc.), ce que Catriona Maccallum appelle « l'ouverture intelligente ».

5.2. Mettre l'édition scientifique au cœur du monde académique, par Ruppert Gatti

De la même manière, Ruppert Gatti dit que les centres de recherche sont très bien placés pour faire de l'édition scientifique. Parmi les nombreux avantages, il note une interaction plus directe avec l'audience recherchée. Il balaie par ailleurs les différents problèmes que l'on pourrait envisager : pour lui, l'expertise nécessaire pour faire de l'édition ainsi que les ressources sont déjà en partie présentes au sein des centres de recherche, et de nombreux outils, souvent en Open Source, sont désormais disponibles pour aider à la mise en place de tels services. Selon lui, le monde académique a tout à gagner à mettre l'édition scientifique au centre de ses activités.

5.3. L'édition universitaire en libre-accès en Europe : difficultés et opportunités dans les sciences humaines, par Victoria Tsoukala

Victoria Tsoukala s'intéresse plus particulièrement aux sciences humaines. La communication scientifique dans ce domaine diffère des STM pour de multiples raisons ; le libre accès, en particulier a du mal à prendre de l'ampleur, alors que paradoxalement de nombreuses initiatives allant dans cette direction existent. Malheureusement, le domaine souffre d'une grande fragmentation et d'une difficulté à atteindre une masse critique et donc une certaine pérennité. C'est pourquoi elle recommande qu'il y ait une meilleure coordination des services et des infrastructures au niveau européen, et que les institutions et les chercheurs eux-mêmes s'investissent d'avantage.

6. Traitement et préservation numérique d'objets scientifiques grands et complexes

Cette session était consacrée aux différents enjeux liés à la manipulation et à l'archivage des données scientifiques.

6.1. Le cycle de vie des données : l'exemple suisse, par Pierre-Yves Burgi(27)

Le projet DLCM(28) met d'abord un accent particulier sur 2 étapes de la gestion du cycle de vie des données pour lesquelles les chercheurs ont besoin de solutions : le traitement et l'analyse des données, appelées aussi la gestion de données actives. Un objet complexe peut être défini comme multidimensionnel : il peut être multi-format, multi-structure, multi-source, multimodal, multi-version. La taille de l'objet contribue aussi à sa complexité. Dans le cadre du projet DLCM, les efforts vont d'abord se porter sur 2 outils(29) : *openBIS* et *SLims* et, pour les humanités digitales, *SALSAH*. Les outils mis en place permettront aux chercheurs de travailler avec leurs données et de les documenter, puis elles seront archivées et le set de données sera référencé avec un DOI. Les challenges à relever sont, le choix d'outils génériques ou spécifiques ; comment rendre le processus intuitif ; et enfin la définition d'un business plan solide et durable.

6.2. La préservation des objets scientifiques complexes : identification et capture des données, par Andreas Rauber

La préservation est avant tout motivée par la réutilisabilité des données. Les versions des logiciels, systèmes d'exploitation, et le matériel informatique altèrent les résultats. Pour assurer la confiance en la répétabilité des données, il est nécessaire de préserver, en plus des résultats, le processus et son implémentation. Les « data management plans » doivent être

remplacés par des « process management plans ». La capture du processus et de son contexte passe par la définition d'un nombre d'informations minimum à capturer de manière automatique, ré-actionnable et réutilisable par une machine. Il est aussi primordial de pouvoir précisément citer le sous-set (dynamique) de données utilisé dans une étude, tel qu'existant à un moment donné. Un groupe de travail de la RDA(30) a émis des recommandations pour rendre citables les données dynamiques, tels que l'horodatage, le *versioning* ou l'attribution d'un identifiant persistant à une requête.

6.3. Le traitement des données de recherche : créer des services de données (ouvertes) au CERN, par Patricia Herterich et Tibor Simko

Au CERN, les efforts se concentrent sur la préservation des données pour les expériences du CERN, l'accessibilité des données à travers le CERN Data Portal et la création de liens vers les autres plateformes utilisées au CERN pour fournir plusieurs points d'accès aux données. Au-delà du problème des petabytes de données, il est essentiel de préserver le processus d'analyse pour pouvoir le reproduire ultérieurement. Un nouveau système, basé sur le logiciel *Invenio*, a été développé (31) : il connecte les outils que les physiciens utilisent pour préserver les données (wiki, logiciels, ou autres) en capturant automatiquement les métadonnées techniques et sémantiques. Concernant les données ouvertes, le CERN Data Portal(32) a été créé afin de rendre des sets de données accessibles au public extérieur, après certaines périodes d'embargo définies selon les politiques des différentes expériences du CERN.

Toutes les initiatives présentées pendant les 3 jours de la conférence montrent que la communication scientifique évolue pour permettre à la science d'avoir plus d'impact et d'être plus efficace. Les différents projets, qu'ils concernent les aspects techniques, les processus ou les politiques, montrent qu'une évolution est en cours vers un nouveau paradigme de la communication scientifique, une communication « ouverte ». Ce nouveau paradigme redonne une place privilégiée au chercheur et à l'institution elle-même.

Les présentations et vidéos des sessions sont disponibles sur le site web de la conférence : <https://indico.cern.ch/event/oai9>

NOTES

- (1) <http://ipython.org/notebook.html> (consulté le 29.10.2015)
- (2) Web de données en français
- (3) <http://linkeddatafragments.org> (consulté le 29.10.2015)
- (4) <https://github.com/LinkedDataFragments> (consulté le 29.10.2015)
- (5) <http://nanopub.org/wordpress/> (consulté le 29.10.2015)
- (6) <http://trustyuri.net/> (consulté le 29.10.2015)
- (7) Kuhn, T. et al. Publishing without Publishers : a Decentralized Approach to Dissemination, Retrieval, and Archiving of Data arXiv:1411.2749
- (8) Kuhn, T[8] Science Bots: a Model for the Future of Scientific Computation? arXiv :1503.04374
- (9) <http://hiberlink.org/> (consulté le 29.10.2015)
- (10) Publications en science, technique et médecine
- (11) <http://robustlinks.mementoweb.org/spec/> (consulté le 29.10.2015)
- (12) <http://archive.is/> (consulté le 29.10.2015)
- (13) <https://openaccessbutton.org/> (consulté le 29.10.2015)
- (14) Accès moyennant paiement
- (15) Open Access
- (16) <http://cofactorscience.com/journal-selector> (consulté le 29.10.2015)
- (17) <http://www.ascb.org/dora/> (consulté le 29.10.2015)
- (18) Clearinghouse for the Open Research of the United States:
<http://www.chorusaccess.org/> (consulté le 29.10.2015)
- (19) <http://dashboard.chorusaccess.org/> (consulté le 29.10.2015)
- (20) Application programming interface
- (21) <http://www.share-research.org/> (consulté le 29.10.2015)
- (22) Modèle "G-E-O" : good, efficient, open
- (23) <https://101innovations.wordpress.com/> (consulté le 29.10.2015)
- (24) www.plosone.org/ (consulté le 29.10.2015)
- (25) <https://www.peerageofscience.org/> (consulté le 29.10.2015)
- (26) <https://publons.com/> (consulté le 29.10.2015)
- (27) Voir aussi l'article de recherche consacré à ce projet dans ce même numéro de RESSI.
- (28) Data Life-Cycle Management
- (29) ELN (Electronic Laboratory Notebooks) / LIMS (Laboratory Information Managing Systems)
- (30) Research Data Alliance
- (31) CERN Analysis Preservation : <http://analysis-preservation.cern.ch/> (consulté le 29.10.2015)
- (32) CERN Open Data portal : <http://opendata.cern.ch/> (consulté le 29.10.2015)