

L'ALPSP 2017, Noordwijk, Pays-Bas

Jour 1. Amsterdam: rencontre avec des collègues et Adrian Stanley, vice-président de « Digital Science ». Le sujet principal était l'initiative à venir de « ReadCube Discovery », initiée par l'ACRS. Également évoqué le projet analytique du lectorat.

Jour 2. Train et bus jusqu'à Noordwijk. Repas en compagnie de Toby Green, éditeur de matériel de l'OCDE, qui vient de publier un article sur l'échec de l'accès libre. Toby a élargi leur performant modèle opérationnel « Freemium » (lire gratuitement, au-delà un paiement est requis). Et il agrège les éditions d'ONG à travers l'OCDE. C'est une avancée très encourageante, puisque la plupart des ONG ne possèdent pas de modèles d'édition viables.

Discussion avec Phil Jones de « Digital Science ». J'ai accepté de lui envoyer une copie de l'étude du projet analytique du lectorat de l'ACRS pour des commentaires et suite à l'intérêt de « Digital Science » de fournir à l'ACRS des services analytiques. Ça a l'air prometteur, ce qui laisse à penser qu'ils pourraient accomplir le travail qu'on attend d'eux et pourraient payer les fondations pour leur valeur ajoutée.

Conférenciers invités: Confiance, vérité et édition savante

Ensemble, ils ont fourni un aperçu général sur l'état des revues savantes dans le contexte des sciences. La première présentation, donnée par un chercheur en sciences, stipulait que le problème majeur auquel sont confrontés les chercheurs et la communication savante n'était pas le plagiat ou les revues prédatrices, mais plutôt l'intégrité de la recherche. Le facteur qu'il a identifié comme étant crucial était l'ambiance générale au sein de la communauté scientifique, qui a été générée, en partie, par les prix discernés aux publications, et plus précisément, le nombre de publications sans récompense nuancée pour faire part de toute la vérité, y compris les échecs et les résultats négatifs. Il a constaté que les chercheurs sont inmanquablement sélectifs dans leurs publications; un comportement qui biaise le bilan scientifique. Son argument était que la solution résiderait dans l'engagement en faveur de « l'intégrité de la recherche » et de la transparence accrue, par le biais de publication de données accessibles. En bref, il a appelé à la refonte du système des récompenses et à l'engagement de tout un ensemble d'acteurs dans la recherche et le processus de la communication savante.

La deuxième conférence principale a été donnée par l'ancien directeur du Bureau de la politique sur les sciences et technologies des É.-U, Kei Koizumi. Son argument principal était que les États-Unis entraient dans une ère semblable à celle que le Canada a connue au cours de ces dernières années. Des données de recherche ont été supprimées, en grande partie, suite aux coupures budgétaires en sciences et des dépenses dans des domaines tels que la Sécurité intérieure. Ainsi, la politique cesse d'être guidée par l'information. Il a utilisé le terme « politique dépourvue de science ». Il a aussi constaté une méfiance croissante à l'égard de l'information scientifique, y compris l'information scientifique sociale, ajoutant que cela s'est intensifié sous l'administration de Trump qui recourt directement à la suppression de l'information incommode et gênante. Ceci nuit à la confiance du public envers la recherche, et cette méfiance prend une forme diffuse et générale, susceptible de réduire la qualité des revues. Son constat général est que l'infrastructure de la recherche est malmenée, mais tient encore, soutenue largement par des projets d'accès public.

Séance plénière no1: Sociétés savantes: Faire face aux grands bouleversements

La première séance plénière s'est penchée sur la façon dont les sociétés savantes font face aux changements liés à la production et à l'exploitation de l'information. L'argument général avancé par les deux intervenants consistait à montrer comment les sociétés savantes modifient leurs rôles traditionnels pour acquérir une place importante dans l'esprit du public. Hetan Shah, directeur

exécutif de la Société mondiale de la statistique, a fait deux observations. Premièrement, le traditionnel soutien de tabouret à trois pattes, regroupant abonnement, rencontres, et publication, régit encore une grande partie des activités de la Société mondiale de la statistique, mais leur approche est considérablement plus innovante. Partant du principe que les sociétés savantes n'ont aucun droit intrinsèque d'exister, il a développé et mis en place un plan stratégique destiné à renforcer la discipline et la profession, accroître la formation statistique au sein du public, et le transfert des statistiques dans l'intérêt public. Il a aussi professionnalisé la compagnie plutôt que de la faire fonctionner largement par des travailleurs bénévoles. En ce qui concerne les revues, il a instauré un contrat de soumission pour les fournisseurs de services. Il fait preuve d'esprit commercial pour générer un financement croisé des abonnements, utilise les membres comme des ambassadeurs statistiques, et implique généralement les membres à la cause de la discipline.

La deuxième conférencière, Alix Vance de « GeoScience World » – un mélange de sociétés axées sur la géologie, qui ont aussi mis l'accent sur une perspective commerciale comme partie intégrante de leur qualité d'éditeur à but non-lucratif, soulignant les économies d'échelle résultant de la collaboration. Elles utilisent une structure commune de risque/récompense pour la publication de leurs revues, livres, et données; elles sont impliqués dans l'innovation des outils de géoscience, et poursuivent la croissance à la fois horizontale et verticale : une bonne combinaison du « fait maison » et de l'édition commerciale.

Finalistes pour les prix de l'innovation sur le site de l'ALPSP

Kudos

Escalex

Publons

Delta Think

INASP Author Aid

Deuxième jour:

Séance plénière no 2: Repenser la publication: les navigateurs, technologies, et financement derrière des évolutions audacieuses

Les panelistes ont abordé « l'économie des articles et l'avons-nous réalisée? » Le sens de cette phrase n'était pas clair, mais en général tous les intervenants s'accordaient à dire qu'alors que beaucoup de changements sont en cours, la publication des articles reste essentielle, car les chercheurs sont récompensés pour le nombre d'articles qu'ils ont publiés et rien d'autre, pas sur les résultats négatifs, pas sur le développement des outils, ni sur la ré-analyse des données, ni sur la reproduction. La renommée, l'impact et les activités qui ne tournent pas autour de l'article ne mènent pas très loin; la bonne vieille tradition de l'évaluation par les pairs est encore de rigueur, la croissance de l'accès libre a stagné, et pendant que l'Occident se transforme les pays en développement sont encore en train d'essayer de rattraper le système traditionnel.

Un conférencier a fait remarquer que les pirates (incluant les éditeurs prédateurs) emboîtent le pas à l'accès libre pour créer un autre modèle d'activité et subvertir l'ensemble du système. Il était constaté qu'une version de PLOS se profile derrière SciHub ainsi que sur son propre site, mais elle ne semble pas vider Plos de son trafic. On a fait remarquer aussi que SciHub est un cul-de-sac, en ce sens que les liens sortants vers d'autres articles disparaissent – ils doivent être copiés et collés. Malgré l'instauration des commentaires des lecteurs, ce système ne semble pas être très utilisé. Des moteurs de prépublications émergent de partout, eg. socRxiv, bioRxiv, mais il n'est pas clair quel

sera leur rôle. Plus tard dans la conférence, on a fait remarquer qu'il n'y avait, en réalité, pas de modèle d'affaire émergent : il y a des frais sans revenus.

Tous comme pour l'accès libre, chacun s'accorde à penser que Science Ouverte est un grand desideratum, mais s'il est fondé sur le modèle traditionnel de l'édition de la revue (c'est à dire gagner de l'argent sur le marché) il n'est pas viable. Encore ici, plus tard dans le programme, un représentant de la Fondation Gates a fait remarquer que non seulement les gens financés par Gates tirent profit de la publication de Science Ouverte Gates (Gates Open Science), qui possède un processus rapide d'évaluation par les pairs et fournit une publication immédiate en accès libre, mais que les projets financés par Gates peuvent avoir suffisamment de fonds pour payer APC.

Implicitement, derrière cette discussion il y avait un appel à une restructuration des mesures incitatives pour que des chercheurs puissent être reconnus pour autre chose que les articles qu'ils publient. En science, au moins, il y a beaucoup d'informations extrêmement importantes non publiées : des données, des outils, des protocoles, à cause de mesures incitatives inadéquates. Un autre appel lancé ici plus tard, encore et encore, est que le financement soit disponible pour couvrir les frais de la publication, comme fait la fondation Gates. Il était rafraichissant de voir que cette discussion n'était pas engloutie par les exigences consuméristes croissantes des bibliothécaires pour des produits gratuits.

On a souligné le besoin de la transparence en rendant les données disponibles. Deux constats communs ressortent de cette discussion. Le premier était le besoin de données analytiques. On recourt davantage aux données analysées, à moins qu'on soit axé sur la question de la reproductibilité, étant donné que la société et les chercheurs sont déjà éprouvés par la surabondance d'informations. Dans une présentation faite par la *New England Journal of Medicine*, il a été constaté que plus de 200 groupes de chercheurs en santé ont participé à un concours de réanalyse de données. Certains ont généré des hypothèses intéressantes pour des recherches ultérieures.

John Inglis a évoqué la fondation de bioRxiv par Cold Spring Harbor Lab et la mise en place d'un transfert facile d'articles depuis ce serveur de pré-impression vers diverses revues. En règle générale, les comités de recrutement et d'accession à la permanence en science, ainsi que les bailleurs de fonds considèrent l'hébergement d'articles dans les serveurs de pré-impression nécessaire, mais pas une démarche suffisante. Plus globalement, le sentiment qui en ressortait était que les serveurs de pré-impression contribuent, en effet, à la valorisation de la recherche en science. Les articles bénéficient de commentaires et des changements sont souvent effectués avant la soumission à la revue. De plus, les chercheurs les consultent quotidiennement pour se tenir au courant de ce qui se passe dans leur domaine. Autres points abordés:

- pré-impression: circulation avant la soumission à l'évaluation par les pairs
- serveur vs service
- serveurs comme mécanisme autonome
- pas de modèle d'affaire pour les serveurs de pré-impressions
- les pré-impressions sont reconnues dans les demandes de subventions et l'accession à la permanence.
- les serveurs de pré-impressions vont-ils être manipulés par des affichages frauduleux pour affaiblir l'influence de la science?

Après cette discussion, je me suis demandé si Calcul Canada serait enclin à fournir un serveur de pré-impression à la recherche en sciences sociales et humaines canadiennes. Liz Allen de F1000, une organisation qui suggère des articles pertinents pour les chercheurs en sciences de la vie, fournit

des outils d'écriture et de recherche et une plateforme pour la recherche scientifique ouverte, a exploré la manière dont les chercheurs peuvent partager davantage ceux qu'ils font, remarquant que la métrique s'articule autour des articles – pas autour de la recherche, comme par exemple les données, les notes, les recommandations, etc. Elle a mentionné le développement d'une plateforme libre accès par F1000 pour l'Institut Neurologique de Montréal.

Hannah de *Wellcome* a fait remarquer que *Wellcome* intervient dans le processus de la communication de la recherche pour remodeler la communication de la recherche, afin d'atteindre ses objectifs et satisfaire ses besoins, qui consistent généralement à rendre la recherche en santé disponible, et accroître la reconnaissance d'autres informations/activités de recherche utilisant la marque *Wellcome* – en somme, assoir sa légitimité.

Kristan Ratan de la Fondation *Collaborative Knowledge CoKo* a parlé en termes d'écosystème de recherche émergent, composé d'une constellation de résultats de recherche, affirmant que, globalement, nous sommes loin d'atteindre un tel bilan de la situation. Pour faire avancer les choses, CoKo a développé KOL un code de documentation à source ouverte pour aider les chercheurs à établir un rapport des activités de recherche autres que la distillation de la recherche en formats d'articles médiocres. WAX, un éditeur html, est l'un de ces outils. Le problème dans le cas présent est que les chercheurs sont concentrés sur les données, pas sur les outils de communication.

Dans une session intitulée « connaître les dernières tendances dans l'industrie », l'auditoire a été familiarisé avec un fournisseur de services pour monographies qui s'appelle GlassTree. Daniel Berse a mis en lumière les services qu'ils fournissent, et qui sont parallèles à ceux fournis aux revues par OJS. Il a beaucoup insisté sur les gains des auteurs et a mentionné les réductions de coût pour les étudiants. Leur devise : ROI ou péris, ne publie pas ou péris.

SAM BRUINSMA de RA21 a décrit une initiative de NISO/STM destinée à venir à bout des obstacles qui empêchent de passer vers une identité fédérée, au lieu de l'adresse d'authentification IP (qui, vraisemblablement, rend le vol possible). Manifestement, quelques produits et services sont disponibles aujourd'hui dans l'espace de découverte d'identité.

Tasha Mellins-Cohen a souligné les activités du projet Cupcake, qui est axé sur le développement de la métrique pour l'évaluation de la qualité des revues savantes (pas en science). Elle vise à créer une base de données sur des informations liées aux services de publications et de ventes, et à communiquer la qualité du service que les éditeurs de la recherche savante ont fournie à leurs clients – à leurs auteurs, bailleurs de fonds, et abonnés.

.....

Andrew Pitts a fait un compte rendu des questions de sécurité en mettant l'accent sur SciHub

Éditorial : Jo Adetunji

*The Conversation* « rassemble journalistes et chercheurs au Royaume-Uni ». Elle est subventionnée par des universités britanniques et est destinée à augmenter l'engagement public afin d'accroître la confiance du public et des médias dans la recherche. Le conférencier a affirmé qu'il y avait une filiale au Canada, mais lorsque j'ai fait ma recherche, la grande partie de ce qui était écrit était plutôt écrit par des journalistes britanniques. D'autre part, les canadiens eux-mêmes ont écrit sur des questions non-canadiennes. La discussion était semblable à un service télégraphique numérique générant du contenu de type « Creative Commons » et contenait un code qui permet de retracer la publication. Elle fournit également des outils d'écriture pour aider les universitaires à écrire pour

le grand public. Ils dirigent aussi une rubrique appelée « Validation des faits » pour déceler des mensonges et des falsifications.

[theconversation.com/uk](http://theconversation.com/uk)

Heather Staines a passé en revue la technologie d'annotation, Hypothes.is. En gros, c'est un logiciel qui permet des couches d'annotations, entre autres, pour le public, pour une communauté d'experts, etc., et ils travaillent à une annotation standard. Leur travail porte globalement sur la capacité des étudiants à évaluer les sources numériques et à promouvoir la compréhension des algorithmes de polarisation numérique.

IA:

Sam Herbert de 67 Bricks a fait une introduction sommaire sur l'IA. Ses exemples d'application à l'édition des revues étaient quelque peu ténus (identifier les évaluateurs, améliorer la communication en marketing (ciblage), mais il a, tout de même, réussi à communiquer à son auditoire les fondamentaux de l'IA. Il a parlé de l'apprentissage automatique, et évoqué comment il peut amener les ordinateurs à identifier le cancer du sein avec une plus grande fiabilité que des personnes formées à ce genre de tâches.

Ruth Pickering de Yewno a décrit comment ils ont créé un logiciel de recherche pour interroger une base de données composée d'un contenu numérique de la bibliothèque de Stanford. C'était vraiment impressionnant. Un étudiant peut simplement taper *The Faerie Queen* et puis s'ouvre à gauche de l'écran le contexte de l'œuvre en question, et ensuite toute la littérature que la bibliothèque possède sur cette œuvre. Un étudiant peut alors créer un nouveau nœud (par exemple, Reine Elizabeth I), et puis un troisième, et voir des informations plus loin qui enrichiraient davantage sa recherche. Il semblerait que le processus de recherche peut mener l'utilisateur directement aux pages précises liées à la requête initiale. Elle a expliqué que ce qu'ils essayaient de faire était d'établir un réseau neuronal. Cela crée le sentiment que le simple fait que cet outil soit disponible permettrait à un étudiant de Stanford de surpasser les étudiants des autres universités.

IA-2 Deux études de cas d'édition

INSPEC est un outil d'indexation de contenus d'ingénierie qui comprend une collection considérable de contenu de revues (peut-être une un ensemble exhaustif). David Smith a montré comment INSPEC avait numérisé sa base de données, puis avait soumis le contenu à des commentaires. Il a décrit les résultats comme étant particuliers, en ce qu'ils étaient exhaustifs ; ils ont recueilli tout le contenu pertinent sur cette enquête. Il a constaté qu'ils n'ont pas utilisé la totalité de l'article, sans dire explicitement si c'était seulement le résumé de l'article ou si c'était le résumé créé par INSPEC.

MARCEL KARNSTEDT-HULPUS a décrit un travail de gestion de bases de données effectué sur le contenu de *Springer Nature*. Il affirme être capable de produire un mode de savoir structuré, fondé sur un océan de faits sémantiquement enrichis et intégrés. Formulé différemment, il a proposé une recherche intelligente qui va au-delà de la recherche par mot-clé et s'est penché sur les intentions des utilisateurs, établissant une cartographie de leurs recherches par des entrées diverses. A l'avenir, il voit ça comme un moyen qui facilite l'exploration interactive du savoir, de la navigation intelligente, et de découvertes inattendues. Il a parlé d'une entité (semi-)automatisée de reconnaissance dans les textes en langue naturelle, des sur-taxonomies, de techniques avancées d'apprentissage automatique, et de données analytiques telles que l'extraction de bases de

données de graphes et de réseaux et les bénéfiques qui en résultent. Enfin, c'est au fruit qu'on jugera l'arbre!

Je n'ai pas assisté à la session sur la perte de Belgian Bro au profit de HP en essayant d'obtenir des droits tarifaires sur les photocopieurs. Heureusement le Canada a largement évité de tels prélèvements, hormis sur les enregistrements audio.

Jour 3

Séance plénière no 4 : Optimiser la valeur des découvertes de recherche et des données : Innovation intercommunautaire.

Carly Strasser de la Fondation Moore (Moore's Law) a décrit les divers projets qu'ils ont soutenus. Elle a insisté sur la découverte axée sur les données, et, de façon peu surprenante, étant donné le caractère appliqué ou concret de la carrière de Moore, l'essentiel de leur soutien sert à promouvoir d'autres activités de recherche que celles liées à la publication d'articles, comme par exemple, le développement d'outils, de codes, etc. Elle a présenté leur projet de développement de logiciel, l'atelier du laboratoire interactif Jupyter qu'ils ont installé sur GitHub (pour qu'il soit téléchargé); ils supervisent les développements sur « Libre Accès » et travaillent à l'estimation de A L (coûts de bioRxiv).

Ashley Farley de la Fondation Gates a parlé de ses activités et de son engagement en faveur du partage ouvert des informations; notez-bien, il ne s'agit pas de Libre accès à proprement parler. Toute recherche financée par Gates est publiée par CC, et est immédiatement disponible (y compris les données). Gates couvre les APC et n'autorise aucune exception, ce qui permet une grande conformité. Cet automne, ils vont lancer ce qui ressemble à une nouvelle opération éditoriale qu'ils n'appelleraient probablement pas revue. Son nom est « Gates Open Access ». Il est possible que ce soit la même chose que le « Gates Open Research » qui avait été annoncé en mars 2017. Je le suppose. Ils comptent publier rapidement, en renvoyant au chercheur un article qui a été soumis au bout de 10 jours, suite à une évaluation ouverte (il n'est pas clair quelles sont les procédures d'évaluation en place, si tenté qu'ils y en aient). Ils ont l'intention de mesurer l'impact et aborder de front les défis de la science ouverte et la gestion des données.

Pam Miller du *New England Journal of Medicine* a décrit le projet précédemment dans ce rapport. Je pense que c'était elle qui a parlé d'écosystèmes scientifiques et de responsabilité.

Marcus Mufano, professeur de psychologie biologique, a examiné les pratiques de recherche assez rigoureusement, mettant en lumière une pratique courante vraiment inacceptable, plus particulièrement celle liée à la production d'hypothèses de recherche après la collecte de données. Plusieurs considèrent cette pratique comme parfaitement acceptable. Ce n'est statistiquement pas fiable. Elle génère des hypothèses plutôt que de les tester. 90 articles en psychologie ont tous confirmé leurs hypothèses. Il a proposé un enregistrement des hypothèses et de la méthodologie préalablement à la collecte, et l'engagement de la revue de publier avant que la collecte n'ait eu lieu. Cela ne va pas avoir lieu. Il a suggéré la création de sceau de qualité pour les revues; il a affirmé que l'ouverture dans l'évaluation, et à travers la publication des données, contribue au contrôle de la qualité; il a recommandé la synchronisation des revues et des organismes subventionnaires avant l'expérimentation; et comme la reproduction est rarement publiée, la science n'est pas très autocorrective.

Séance plénière 5. IA Les robots arrivent: présidente, Isabel Thompson

Joanna Bryson, une professeure de Bath, a affirmé que la viabilité et l'inégalité sont les questions essentielles, pas les robots. Elle a également plaidé en faveur de la responsabilité humaine dans chaque action entreprise par les robots, ou, pour ainsi dire, n'importe quelle autre IA ou algorithme.

Sebastian Huempfer de Echobox a décrit sa compagnie comme un fournisseur d'automatisation intelligente. Cette société distille les données et conçoit des stratégies pour l'identification des auditoires. Elle fournit des manchettes pour divers sites d'informations en ligne, tels que *The Guardian*, qui guide les lecteurs vers le contenu principal. Elle a aussi des consultants des media sociaux.

Venture Capitalist. Volker Hirsch a survolé une série de développements alarmants liés au remplacement des travailleurs humains par des robots (et de scenarios semblables déconcertants). À titre d'exemple, la Chine a remplacé par des robots 60 000 travailleurs qui gagnaient 20 dollars par jour. Il a remarqué que les données accélèrent l'IA, car elle fournit les fondations à l'IA. Des robots sont maintenant en train de former des robots. Le programme Deepface est plus efficace que les humains dans le domaine de la reconnaissance faciale. Les machines peuvent inventer des langues qui fonctionnent pour eux, mais pas pour les humains. Des avocats sont remplacés par l'IA; tout ce qui touche à la reconnaissance des formes est probablement mieux géré par la machine. Les présentations ont été suivies de discussions animées.