

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Résumé

Dans le cadre de ce projet, des données d'utilisation liées à des bibliothécaires universitaires seront collectées à partir de quatre sites web de profils universitaires, soit Google Scholar Citations, ResearchGate, Academia.edu et ORCID (Open Researcher and Contributor ID). Ce plan de gestion des données décrit le processus de collecte de données, les besoins en matière de stockage et de conservation des données à long terme, ainsi que les enjeux liés aux données qui peuvent surgir au cours du processus de recherche.

Détails administratifs

Nom du projet :

Étude sur l'utilisation des sites web de profils universitaires par les chercheurs bibliothécaires

Chercheuse principale/intendante des données de recherche :

Maha Kumaran et Li Zhang

Responsable des données du projet :

maha.kumaran@usask.ca; li.zhang@usask.ca

Description:

Les sites web de profils universitaires sont des plateformes en ligne où les chercheurs créent leurs profils professionnels, et organisent et présentent leurs résultats de recherche. Parmi les exemples de sites web de profils universitaires, il y a Google Scholar Citations, ResearchGate, Academia.edu et ORCID (Open Researcher and Contributor ID). Au cours des dernières années, on observe une utilisation croissante de ces sites web parmi les chercheurs, principalement en raison de leur valeur perçue pour faciliter la collaboration et augmenter l'effet de la recherche. Alors que les bibliothécaires aident les chercheurs à développer des profils sur ces plateformes, on ignore comment les bibliothécaires eux-mêmes, en tant que chercheurs, utilisent ces outils. Ce projet étudiera la présence de chercheurs bibliothécaires sur les sites web de profils universitaires (c'est-à-dire les bibliothécaires qui ont l'obligation professionnelle de mener des recherches) aux États-Unis et au Canada. Les résultats de ce projet permettront de mieux comprendre les sites web de profils universitaires, de trouver de nouvelles orientations pour améliorer les fonctionnalités de ces sites et de découvrir les possibilités de fournir de nouveaux services de bibliothèque à la communauté des chercheurs.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Collecte de données

Quels types de données recueillerez-vous, créez-vous, couplerez-vous, acquérez-vous ou consignerez-vous?

Des données textuelles et numériques seront recueillies dans le cadre de ce projet. Les données textuelles spécifiques comprennent les noms des chercheurs bibliothécaires, leur rang et leur établissement, leurs affectations et les questions posées et répondues sur les sites web des profils universitaires. Les données numériques collectées comprendront les publications répertoriées, les lectures, les opinions, les suiveurs, les abonnés, les citations, l'indice h et la cote RG (une mesure fournie par ResearchGate). Étant donné que les quatre sites web de profils universitaires ne fournissent pas tous les mêmes informations, les données énumérées ci-dessus ne s'appliquent qu'à certains sites web.

Nous choisirons des chercheurs bibliothécaires universitaires au Canada et aux États-Unis comme objets d'étude. Ces chercheurs bibliothécaires doivent provenir d'établissements membres de l'Association des bibliothèques de recherche du Canada (ABRC) ou de l'Association of Research Libraries (ARL) qui exigent qu'ils publient pour obtenir un poste permanent, une promotion ou une nomination continue. Les sites web de profil académique seront choisis parmi ResearchGate, Academia.edu, Google Scholar Citations et ORCID — les quatre principaux sites web de profils universitaires utilisés par les chercheurs bibliothécaires au Canada et aux États-Unis. Les informations sur les chercheurs bibliothécaires universitaires seront recueillies à partir des sites web de l'ABRC et de l'ARL.

Selon quels formats de fichiers les données seront-elles recueillies? Ces formats permettront-ils la réutilisation, le partage et l'accès à long terme aux données?

Un tableur MS Excel et une base de données MS Access seront utilisés pour recueillir les informations. Les données de la base de données seront transférées à SPSS pour analyse. Une fois le projet terminé, nous convertirons le format des données en un format non-propriétaire, tel que le format CSV (valeurs séparées par virgules), afin de faciliter la réutilisation, le partage et la conservation des données.

Les formats de fichiers propriétaires (par exemple, MS Excel, MS Access) sont associés à un logiciel particulier, qui peut ne pas être disponible à long terme. Il est donc préférable d'enregistrer les fichiers dans un format non-propriétaire (par exemple, le format CSV pour les bases de données et les données tabulaires) pour les conserver à long terme.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Quelles conventions et procédures utiliserez-vous pour structurer et nommer vos fichiers, et pour gérer leur version, afin de vous aider, ainsi que les autres, à mieux comprendre comment vos données sont organisées?

Chaque fichier sera nommé avec une courte description/abréviation reflétant son contenu, suivi de la date de création. Pour enregistrer les différentes versions, nous ajouterons un numéro de version dans le nom du fichier. Par exemple, le nom de fichier GSC_20200608_v01.xls représente les données sur les Google Scholar Citations acquises le 8 juin 2020, la 1^{re} version.

Nous allons créer un document pour préciser les conventions de nomenclature des fichiers et fournir une liste d'explications sur les courtes descriptions/abréviations utilisées dans les noms de fichiers.

Documentation et métadonnées

Quels seront les documents nécessaires pour que les données soient lues et interprétées correctement à l'avenir?

Nous fournirons une brève description du projet, une méthodologie détaillée sur la manière dont les données sont collectées, la date de la collecte des données, l'analyse effectuée et les détails sur les personnes qui ont effectué chaque tâche pour accompagner les données.

Un dictionnaire de données sera créé pour définir les différents champs de données, le type de données et l'explication du codage des données.

Comment vous assurez-vous que la documentation est créée ou saisie de façon uniforme tout au long de votre projet?

Les chercheurs se réuniront régulièrement pour discuter des questions ou des problèmes qui pourraient survenir au cours du processus de recherche : par exemple, comment traiter les variations de nom et comment traiter les données inexactes ou incomplètes sur les sites web de profils universitaires. Toutes les décisions prises seront consignées dans le document de méthodologie afin de s'assurer qu'elles tiennent compte des changements intervenus dans le processus de collecte des données.

La documentation des décisions et des modifications apportées au processus de collecte des données vous permettra également de gagner du temps lors de la rédaction ultérieure du manuscrit.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Si vous utilisez une norme ou des outils de métadonnées pour documenter et décrire vos données, veuillez les énumérer ici.

Les [schémas de métadonnées Dublin Core](#) seront utilisés pour documenter et décrire les données. Il s'agit d'un des schémas de métadonnées les plus utilisés dans les dépôts disciplinaires et institutionnels ; ils comprennent 15 éléments de métadonnées. Les principaux éléments du Dublin Core que nous envisageons d'utiliser comprennent le titre, le sujet, la description, le créateur, la date, le type, le format et l'identifiant (par exemple, l'identifiant d'objet numérique [DOI]).

Stockage et sauvegarde

Quels sont les besoins de stockage prévus pour le projet (en mégaoctets, gigaoctets, téraoctets, etc.) et quelle sera la durée de stockage?

Selon les données recueillies dans une étude précédente portant sur les chercheurs dans d'autres domaines, nous avons estimé que la taille globale des données brutes pour ce projet sera de 20 à 50 mégaoctets. Cela ne tient pas compte des multiples versions de fichiers. Le besoin final de stockage sera d'environ 500 mégaoctets lorsque plusieurs versions de fichiers sont prises en compte.

Vous devrez souvent sauvegarder plusieurs versions des données collectées. Les besoins de stockage des données seront donc au moins plusieurs fois supérieurs à la taille des données brutes.

Comme les sites web de profils universitaires changent rapidement, nous envisageons de conserver les données acquises, y compris les données brutes et celles générées à partir des données brutes dans l'espace [OneDrive](#) pendant au moins 7 ans, conformément aux exigences des bailleurs de fonds communs de la recherche.

Notez que la durée de stockage correspond à la durée pendant laquelle vous avez l'intention de conserver les données sur votre propre espace de stockage (OneDrive pour ce projet) ; cette durée diffère de la conservation à long terme dans un dépôt de données.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Comment et où vos données seront-elles stockées et sauvegardées pendant votre projet de recherche?

Les données seront stockées dans l'espace [OneDrive](#) institutionnel de l'Université de Saskatchewan, soit un stockage infonuagique sécurisé développé par Microsoft. Il est facile d'y accéder à partir de différents types d'appareils et peut être partagé par l'équipe de recherche. De plus, il est sauvegardé automatiquement pour éviter toute perte de données.

Comment l'équipe de recherche et les autres collaborateurs pourront-ils accéder aux données, les modifier et y contribuer tout au long du projet?

L'accès aux données de recherche sur OneDrive sera limité aux membres de l'équipe de recherche. L'utilisation de OneDrive minimise les efforts nécessaires au transfert des données. Nous prévoyons que le travail initial de collecte des données sera réparti entre les deux chercheuses. Lorsque le travail initial de collecte de données sera terminé, les données seront vérifiées par l'autre chercheuse pour s'assurer de leur exactitude et de leur intégralité. Après l'inspection finale, les données collectées par chaque chercheuse seront regroupées dans un fichier principal.

Conservation

Où déposerez-vous vos données pour la conservation à long terme et l'accès à la fin de votre projet de recherche?

Bien que notre établissement ne dispose pas actuellement d'un dépôt de données, nous envisageons de déposer les données acquises dans un dépôt de données canadien ouvert qui accepte les données d'autres établissements tel que [Dataverse de l'Université de l'Alberta](#). Dataverse est une application web de dépôt de données à code source ouvert permettant de déposer, partager, conserver, découvrir et citer des données de recherche. Les données déposées dans Dataverse de l'Université de l'Alberta peuvent être découvertes à partir d'autres plateformes de dépôt telles que le DFDR (Dépôt fédéré de données de recherche), ce qui augmente la possibilité de découvrir nos données.

Pour en savoir plus sur les dépôts de données, communiquez votre bibliothécaire pour en trouver un qui convient à votre discipline ou à votre établissement.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Indiquez comment vous vous assurerez que vos données soient prêtes à être conservée. Envisagez des formats de fichiers favorables à la conservation, en assurant l'intégrité des fichiers et la dépersonnalisation, et l'inclusion des documents justificatifs.

Comme les données collectées seront en formats MS Excel et MS Access, nous les convertirons en format CSV pour les conserver à long terme. Les documents d'accompagnement seront convertis en format TXT pour aider les futurs chercheurs à comprendre et à réutiliser les données.

Bien que cette étude ne comprenne pas de données sensibles puisque toutes les informations collectées seront accessibles au public, nous rendrons anonymes les noms des sujets de l'étude, car nous ne croyons pas que ces informations intéresseront les futurs utilisateurs des données.

Partage et réutilisation

Quelles données partagerez-vous et sous quelle forme les partagerez-vous (p. ex. brute, traitée, analysée ou définitive)?

Les données traitées seront partagées après avoir dépersonnalisé les noms des objets de l'étude. Les données seront partagées en format CSV.

Le partage des données brutes n'est pas toujours nécessaire pour diverses raisons (par exemple, confidentialité, taille des données brutes). Réfléchissez à la manière dont les futurs chercheurs pourraient utiliser vos données et choisissez le format pour le partage des données en conséquence.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Avez-vous envisagé le type de licence d'utilisateur final à inclure dans vos données?

Nous aurons la licence Creative Commons Attribution CC BY pour les données, qui permet à d'autres personnes de distribuer, réutiliser, adapter et développer les données à condition que les créateurs des données originales soient crédités.

Quelles mesures seront prises pour aide à faire connaître vos données dans le milieu de la recherche?

Nous présenterons le projet lors de conférences, publierons les résultats dans une revue révisée par des pairs avec des liens vers les données, et communiquerons l'existence des données par le biais des médias sociaux et du bouche-à-oreille. Nous espérons qu'un identifiant d'objet numérique (DOI) sera attribué aux données. Un identifiant d'objet numérique fournit une identification permanente et persistante aux données auxquelles il est associé, rendant ainsi les données plus facilement accessibles aux moteurs de recherche et aux autres chercheurs.

La citation des données déposées sera incluse dans la publication de la revue, et la citation de la publication sera incluse dans les données déposées afin que les informations puissent être découvertes des deux manières.

Responsabilités et ressources

Indiquez qui sera responsable de la gestion des données de ce projet pendant et après le projet, et les principales tâches de gestion des données dont cette personne sera responsable.

Les responsabilités en matière de gestion des données seront partagées également entre les chercheuses en raison de l'envergure relativement réduite du projet. Les principales tâches de gestion des données sont les suivantes : identification des sujets de l'étude, élaboration de la documentation, création d'une base de données MS pour la collecte des données, collecte des données sur quatre sites web de profils universitaires, inspection de l'exactitude et de l'intégralité des données, analyse des données, conversion des données et versement des données dans un dépôt de données ouvert.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Comment les responsabilités relatives à la gestion des activités de données seront-elles gérées si des changements importants surviennent dans le personnel qui supervise les données du projet, y compris un changement de chercheur principal?

Si l'une des principales chercheuses n'est plus en mesure de participer au projet, l'autre principale chercheuse est censée poursuivre le projet. Dans ce cas, on peut demander au laboratoire de recherche en sciences sociales (SSRL) de l'établissement d'élaborer un script permettant de collecter automatiquement des données sur les sites web de profils universitaires, bien que les données collectées par cette approche puissent ne pas être aussi fiables que la collecte manuelle.

Une autre solution consiste à mener la recherche à plus petite échelle, par exemple en étudiant les chercheurs bibliothécaires au Canada uniquement, ou en sélectionnant deux sites web de profils universitaires au lieu de quatre.

De quelles ressources aurez-vous besoin pour mettre en œuvre votre plan de gestion des données? À combien estimez-vous le coût global de la gestion des données?

Le stockage des données (OneDrive) pendant la phase active du projet est fourni par l'établissement ; il n'y a pas de coût supplémentaire. Comme nous envisageons de déposer les données dans un dépôt de données ouvert, nous ne prévoyons pas non plus de coûts directs pour la conservation des données à long terme.

Si nous devons utiliser le SSRL pour développer un script pour la collecte de données, le coût prévu serait d'environ 2 000 dollars. Dans ce cas, nous demanderons un financement de la recherche pour soutenir le travail. Autrement, nous avons également accès à d'autres sources de financement que nous pouvons utiliser pour ce projet.

Songez à l'avance aux ressources ou aux budgets supplémentaires qui pourraient être nécessaires pour la gestion des données.

Plan de gestion des données pour Utilisation des sites web de profils universitaires (v. 1.0)

Conformité aux lois et à l'éthique

Si votre projet de recherche comprend des données sensibles, comment veillerez-vous à ce qu'elles soient gérées de façon sécuritaire et accessible uniquement aux membres approuvés du projet?

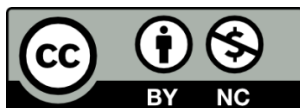
Puisque toutes les informations nécessaires à ce projet sont accessibles publiquement, le projet ne comporte aucune donnée sensible.

Quelles stratégies entreprendrez-vous pour aborder les utilisations secondaires des données sensibles, le cas échéant?

Bien que ce projet ne comporte pas de données sensibles, nous aurons les noms, le rang et les informations sur l'établissement des sujets de l'étude. La publication des noms des personnes peut causer des soucis. Par conséquent, nous remplacerons les noms par des numéros de sujet d'étude, tel que chercheur bibliothécaire 1, 2, 3, etc., lors du versement des données dans un dépôt à la fin du projet.

Comment allez-vous gérer les questions juridiques, éthiques, et de propriété intellectuelle?

Comme cette étude ne contient pas de données sensibles, nous ne pensons pas qu'il y aura des problèmes d'ordre juridique, éthique ou de propriété intellectuelle.



Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
Portage Network | portage@carl-abrc.ca | portagenetwork.ca/fr