Publier un data paper

Un data paper est un article scientifique validé par les pairs qui a pour but de décrire un jeu ou des jeux de données.

Produit de publication de données, pouvant apparaître dans un data journal ou dans une revue académique classique. Contrairement aux articles de recherche classiques, les data papers ont pour but de rendre les données accessibles, interprétables et réutilisables, plutôt que de tester des hypothèses ou présenter de nouvelles analyses.

Quels sont les enjeux de ces publications ?

- Stocker de manière pérenne les données
- Valoriser et faire connaître les jeux de données, ce qui en facilite la réutilisation.
- Communiquer sur l'existence des données et permettre de les trouver ;
- Créditer les auteurs (reconnaissance, référence citable) et valoriser les données ;
- Faciliter la réutilisation des données (en les rendant intelligibles).

Transcription textuelle de la vidéo

C'est un genre d'articles scientifiques qui a plutôt aidé développer jusqu'à maintenant dans domaine sont durs et qui commence à s'implanter dans ce domaine des articles qui vise d'abord à rendre compte d'un jeu de données la manière dont le jeu de données a été créé et qui présente un certain nombre d'informations contextuelles sur la méthode si binaire à travers laquelle là je donnais a été produit et qui va proposer l'information sur les manières de manipuler les données et sur les manières de les réutiliser fait une information surtout descriptif la relation entre les data paper.

Explications complémentaires : le data paper va permettre de répondre à ce manque en fait que nous pouvons trouver dans les articles scientifiques plus traditionnelle pour justement aller plus loin dans la description la partie mythologiques dans le domaine de la biodiversité on a une typologie principal qui est le data paper ou très descriptif dans le jeu de données d'observation de collection il y a de naughty de data paper qui sont plus en sciences humaines et sociales ou là le chercheur a véritablement un peu plus ce décret lille source des données et les choix de construction cela peut aller jusqu'à un véritable article de recherche dans lequel les auteurs vont également décrire la problématique faire un état de la question avec beaucoup de références il ya des conception à géométrie variable ce que doit être un data paper une des bonnes pratiques recommandées c'est de proposer un plan de gestion des données pour détailler donc étape par étape comment les données vont être collectées comment elle va être hébergé stocker numériser les fois qu'on a chez un an mais le circuit logique finalement c'est d'en faire un allié.

Ceci implique que les données soient déposés sur une plateforme notamment entre la paix les entrepôts de données de recherche les entrepôts propose une méta données qui sont déjà faire qui répondent aux critères dabrin corps qu'il faut remplir comme le titre les auteurs les résumés peuvent y être associés également des vocabulaires contrôlée qui représente une source de mots clés ou de concepts admis par la communauté ces éléments là qui sont primordiaux pour pouvoir réutiliser les jeux de données ça va permettre en fait au site internet spécialisé dans les données de moissonner automatiquement ses métadonnées si je mets france troisième république ah oui voilà c'est le premier donc on tombe sur le lien qui est dans data gov le fait de mettre à disposition des jeux de données et des articles de donner les qualifiant ainsi que le chercheur à réfléchir à sa façon de publier le fonds des articles avec une structure relativement rigide un modèle un template est souvent proposé aux auteurs l'idée c'est de commencer évidemment par des éléments de contexte pour bien contextualiser ces données ensuite décrire très précisément la méthodologie de construction des données dans un troisième temps procédé description extrêmement précise du contenu de la base de données donc les différentes variables nous manquait les entrepôts elle se situe et enfin de procéder à une validation technique des données dans un dernier temps je pense qu'il est toujours bien de donner un petit exemple de réutilisation de ces vols et après il ne s'agit jamais que d'une trame et ça convient pas à tous les types d'énergie nous avons eu ensemble de critères ceux qui ont trait à la base de données proprement dit est ce que elle constitue un apport véritablement original il est important aussi que il y à une adéquation entre ce que l'auteur indique vouloir faire en termes d'utilisation de ces données pour la recherche et les résultats qu'ils en tirent nous sommes aussi très attentifs aux questions de licence donc de re distribution de partager des données les évaluateurs peuvent être amenés à vérifier des données intermédiaires qui vous permet la production des données finales les relecteurs vont regarder les données et voir si elles sont bien compatibles avec avec les standards recommandés par la communauté scientifique. Le jeu de données va être publié en même temps que data papers il va obtenir un des eaux y à un identifiant unique qui va être une reconnaissance de sa validation et de sa publication ce sont des articles évolutif aussi donc il y a régulièrement des possibilités de mise à jour à la fois des articles eux mêmes mais aussi pour les jeux de données c'est un signe envoyé aux générations futures à tout moment leur travail peut être enrichie consolidé par les générations suivantes

Transcription textuelle de la vidéo

le rôle des data paper dans la recherche actuellement est en émergence il incite à davantage de transparence à faire un pas vers la science ouverte et faire un autre pas vers la reproductibilité des expériences aujourd'hui le role des data paper dans la recherche à mon sens il est vraiment d'accompagner la question de l'ouverture des données tous autant qu'on est à chaque fois qu'on fait un projet numérique met au point des méthodes et on a besoin de partager ses méthodes le spécifiquement ce data paper est un outil qui va pouvoir être utilisé par les chercheurs donc en fait ça donne une nomenclature commune pour la recherche sur la troisième république l'idée c'était de permettre aux chercheurs travaillant sur cette période de l'histoire française avoir un référentiel commun et de mettre en commun leurs projets de recherche on s'adresse à la communauté scientifique en général à la fois des chercheurs mais aussi beaucoup de plus en plus la communauté académique ce qui est aussi un surnom clients aussi des étudiants et le data piper est l'un des maillons pédagogique de cette chaîne main qui permet de montrer très très précisément comment un processus de recherche est passé de soupçons aux données aux traitements sur les données à l'analyse scientifique.

Les journaux qui publient des dettes à paper on a lectorat qui pussent transdisciplinaire que les journaux de recherche précis qu'il ya un effort à faire pour donner une certaine dose de transdisciplinarité on va taper par il faut faire d'un effort de traduction de certains nombres baie de controverses théorique ou de d'explicitation de concept au d'explicitation de traduction de jargon 6 pin up.

Si le data paper n'est pas encore reconnue à sa juste valeur c'est essentiellement parce qu'il n'est pas encore admis ou intégrés à une chaîne de production académique classique fait affaire sur la reconnaissance de la publication des dettes abord parce que si on veut encourager ce soit ça va passer donc il va être crucial je pense pour les comités de recrutement et les comités d'évaluation de valoriser ce genre d'articles et de valoriser le travail des producteurs de données qui en fait produisent un bien public pure qui est la donne et documentée je pense que l'enjeu ce point de vue c'est en sorte que le data paper puissent être conçues pratique est reconnue comme étant vrais articles scientifiques concernant l'impact des datas paper dans le monde de la recherche cette impacterait est encore sous-estimé devrait être considéré au même titre qu'un article scientifique comme fondamental également permettre la réutilisation et donc limiter le financement d'autres projets par meilleure connaissance de ce qui existe déjà accessible à tous c'est que nous avons constaté c'est que là où il ya des données et là où il y a n'avait pas les données sont davantage jusqu'à trois fois plus que les charges et qu'elle est donnée pour lesquels il n'y a pas de data fait le data paper va permettre une citation plus facile des données réutiliser le génie femme également de tracer des citations de ce jeu de données dans d'autres articles scientifiques on le voit ici grâce aux boutons citation c'est une façon de garder son nom attaché à une information qui reste présente et réutilisables de manière pérenne.

Otg biffle et premier data paper remonte à 2011 2012 maintenant on a pu contacter leurs auteurs qui nous en dit que la publication du data paper avait conduit à

beaucoup de prise de contact et de collaboration de chercheurs français ou étrangers qui voulait consulter les collections dont il avait avant que jamais entendu parler je crois que la dette avait perfidement des cuirs et à ce qui va être important également par la suite à celles exécutées gold bugs et c'est effectivement cette question d'écosystèmes de la publication scientifique publiée en shs c'est beaucoup plus divers et quelque part beaucoup plus compliqué mais aussi beaucoup plus enthousiasmant ceux qui écrivent des datas paper aujourd'hui vont voir les bénéfices de cette pratique dans cinq ou dix ans et on verra je pense en disant ce qu'on saute et d'entrain aujourd'hui et ce qui me passait aujourd'hui

Structurer un data paper

La structure du data paper varie selon les revues scientifiques qui proposent des modèles de fichiers, de présentation ou d'organisation (templates).

La caractéristique du data paper est qu'il est lié au jeu de données brutes qu'il décrit. Le jeu de données est déposé dans un entrepôt de données, au préalable ou au moment de la soumission de l'article. L'identifiant pérenne du jeu de données est indiqué dans le data paper.L'accessibilité des données est vérifiée par les pairs lors de la révision du manuscrit (peer-reviewing). La plupart des revues préconisent des entrepôts de données de confiance.

Transcription textuelle de la vidéo

Pour publier un data paper il faut avant tout choisir sa revue.
Vous pouvez choisir des revues dédiées à ce type d'article, les datas journals, ou bien des revues classiques qui en publient aussi.
Vérifiez bien cependant qu'elle effectue le peer-review du data paper.
Trouvez ensuite l'entrepôt où déposer vos données, en sachant que l'éditeur peut déjà vous en recommander un.
Déposez les et, si l'entrepôt ne s'en charge pas automatiquement, attribuez-leur un identifiant pérenne du type DOI.
Rédigez ensuite votre article est la cover letter associée.
Pour l'article, suivez les instructions aux auteurs, les modèles mis à disposition ou appuyez-vous sur des articles déjà publiés.
Certaines revues disposent aussi d'outil de rédaction en ligne, ou demandent de soumettre le data paper depuis des plateformes externes.
N'oubliez pas de lier votre article à l'entrepôt, en insérant le DOI du jeu de données.
Le comité de lecture aura en effet besoin d'accéder à vos données pour pouvoir évaluer votre data paper.
Vous pouvez maintenant soumettre votre article. Le jury va l'évaluer
selon différents critères, comme l'importance et l'originalité des données ou encore leur accessibilité.
S'il n'y a pas de corrections à apporter, le data paper est validé et il ne reste plus qu'à lui attribuer un identifiant à son tour.
Il devient alors citable et un lien réciproque avec les données déposées dans l'entrepôt peut s'établir.
Notez enfin que si vos données sont directement intégrées dans le data paper, les étapes liées à l'entrepôt deviennent facultatives.

Transcription textuelle de la vidéo

Il existe un certain nombre d'entrepôts qui sont généralistes et accessible à tous donc je pense par exemple ici à zen'auto ou par exemple n'a qu'à la 2e ou magnum qui est plutôt dirigés vers les sciences humaines et sociales ou enfin le harvard à la baisse l'entrepôt doit être à la fois accessible et en même temps sécurisé et puis surtout et rennes ont reprendre l'exemple d'un data paper en archéologie nous avons à cet endroit là l'endroit où est déposée le jeu de données avec le diy cliquable associés qui permet de renvoyer directement aux jeux de données sur la plateforme menlow un outil intégré publication peut permettre de faciliter à la fois la préparation au dépôt des jeux de données en entrepôt et la constitution du data paper dans le domaine de la biodiversité le g beef a mis en place un intégrité publication toul donc je vais vous montrer une ressource qui a déjà été publié sur liberté du g10 france on arrive sur la page du jeu donnait donc c'est un jeu de données qui contient un peu plus de 151 me donner c'est ici qu'on a chargé les données sources et donc quand on clique sur élite on voit les métadonnées qui ont été complétés et vous voyez ici qu'on a plusieurs autres sections plus on aura des méta données détaillées dans plus ça sera facile de rédiger le data paper.

Il ya deux façons d'écrire un data paper donc soit on l'écrit nous mêmes soit on fait appel effectivement une extraction automatique des métadonnées lancer l'outil a refleuri team tout dépend ce que je vais vous montrer donc on va choisir le journal dans lequel on souhaite soumettre le data paper

En cliquant sur une porte manuscrite qui va nous intéresser c'est la première une porte fourni mails metadata le logiciel dont gouttière fin a déjà identifié la liste des auteurs grâce aux métadonnées et dans la voile à l'outil qui a interprété les métadonnées les affiches dans le temple est de pannes soft donc là vous vous retrouvez sur la partie gauche de la page toutes les sections dans lesquels naviguer ou pourront rédiger votre propre data paper sas et hop il existe mais à titre expérimental une chose cependant qui pourrait être sans doute assez facile d'imaginer c'est que lorsqu'on dépose un jeu donné sur un entrepôt il y a tout de suite une petite alerte chimique deans est ce que je ne pensais pas qu'il faudrait accompagner ce jeu de nuit par un data paper.

Nous avons besoin d'entrepôts qui puisse nous permettre de développer des templates spécifiques et adaptées à la nature des données que nous publions l'idéal serait évidemment que ses entrepôts puisse être directement connectée à notre application de gestion de manuscrits.

Qui dit d'attaquer peut dit aussi très souvent modification du jogo donc une modification en suite de l'analyse est donc mise à jour éventuellement d'un article pour moi aujourd'hui le fait de ne pas pouvoir faire de versionning sur de la publication scientifique me semble antinomique en fait avec ce type de pratiques.

Une des difficultés qu'on rencontre aussi c'est de pouvoir transmettre les jeux de données de manière anonyme ise alors anonymisées tant du point de vue des auteurs mais aussi du point de vue des évaluateurs qu'ils ne doivent pas apparaître dans des pots de leur évaluation le data pays port reste très marginal par rapport à la publication scientifique et une récente étude a montré que ça représentait moins d'un pour cent je crois la publication scientifique global n'existe pas en France de data journal dédié aux sciences sociales et en Europe il existe un seul intérêt journal généraliste.

Une copie pure et simple du data paper et de ces informations seraient extrêmement repérables j'espère que une méfiance est en train de se construire par me l'univers de chercheurs à l'encontre de ces revues prédatrice qui sollicitent les chercheurs continuellement chaque jour je reçois entre 20 et 30 demandes de publication de cette nature.

Lorsqu'on regarde les articles ont été publiés à 10 ans a très très grand nombre d'entre eux ne sont pas réplicable les données sont accessibles en ligne l'obstacle principal c'est le portage est donné donc là il y a un problème de culture qui doit être dépassée de partage des données se fait en grande partie pour le moment en dehors des data paper vous écris un data paper il ya beaucoup de compétences qui sont nécessaires soit il faut que cherchait qu il soit qu'ils aient la possibilité de travailler en Haïti moi j'ai fait de vraiment pourquoi des ingénieurs d'études énergie vient de recherche permanent pour pouvoir prendre en charge justement j'ai pas la question de la sensibilisation la question la formation la question aussi du temps que ça que ça représente.

Pour pouvoir unifier les points de vue de ses différents acteurs qui renvoie finalement à des métiers dr professionnelle assez différente ces difficultés je pense que la vraie question justement ça va être de mettre en commun ses différentes expertises pour rédiger le data paix par des chercheurs qui produisent des données ont maintenant une incitation à avoir une meilleure qualité de travail et à mieux décrire leurs données on imagine bien que les data papers soient appelés à jouer un rôle de plus en plus important dans cet échange se partage des informations

Contact
Pour toute question sur les données de recherche
guichet-ardoise@groupes.renater.fr
Florence Thiault

Co-responsable de l'URFIST Bretagne-Pays de la Loire

florence.thiault@univ-rennes2.fr

Aller plus loin

- Le contenu d'un data paper