

Zenodo, Figshare, etc.: rechercher et partager des données de recherche grâce aux entrepôts de données

Atelier des BU - Sciences humaines



Ce contenu est mis à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 3.0 France](https://creativecommons.org/licenses/by-sa/3.0/fr/).

Programme

La mise à disposition des données associées aux articles ou données sous-jacentes (underlying data) est une demande croissante de la part des éditeurs et des agences de financement. Les entrepôts de données permettent de répondre à ces exigences, en assurant l'accessibilité, l'archivage et/ou la diffusion des données déposées, en conformité avec les principes FAIR. Cette session présente les différents types d'entrepôts de données et des recommandations pour identifier un entrepôt adapté.

→ Introduction

→ Trouver un entrepôt de données généraliste

→ Trouver un entrepôt de données spécialisé



Introduction

Terminologie : données de recherche

Modes de partage des données de recherche

Enjeux

Entrepôts de données et principes FAIR

Terminologie : données de recherche

→ Définition large OCDE

Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.

→ Définition NIH : [NOT-OD-21-013: Final NIH Policy for Data Management and Sharing](#)

“Scientific Data: The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.”

→ Un sous-ensemble : les données sous-jacentes à une publication ou *underlying data*

- › Y compris la documentation et le code informatique associés à ces données

Modes de partage des données de recherche

→ Répondre manuellement à des demandes



Exemple : <https://doi.org/10.1186/s12888-020-03010-3>

→ Annexer à un article des fichiers de *Supplementary data*, publiés sur le site de l'éditeur en même temps que l'article

→ Déposer les données dans un entrepôt de données



Exemple : <https://doi.org/10.5255/UKDA-SN-852431>

Availability of data and materials

Readers and all interested researchers may contact Runsen Chen (Email address: runsen.chen@psych.ox.ac.uk) for details.

UK Data Service
ReShare



Home Legal Review procedures Metrics

Emotional intelligence in social work

Biggart, Laura Paule (2018). *Emotional intelligence in social work*. [Data Collection]. Colchester, Essex: UK Data Archive. [10.5255/UKDA-SN-852431](https://doi.org/10.5255/UKDA-SN-852431)

This interdisciplinary project examines whether emotional intelligence helps professionals in emotionally demanding jobs such as social work to perform well and whether it helps them to manage their anxiety and stress. Emotional intelligence indicates how good someone is at understanding how emotions work, recognising emotions in themselves, recognising emotions in others, managing their own emotions and managing emotions in others. This project aims to assess the influence of Emotional Intelligence (EI) training for social workers in relation to burnout and performance using a randomised control trial. It also aims to produce a psychometric measure of social work performance created from the perspective of social workers, managers, service users and educators. The three year study has four phases: (1) a literature review and focus groups will identify items which tap into behavioural characteristics of social worker performance; (2) will validate a social work performance measure created from phase 1; (3) an emotional intelligence intervention will be designed using elements from existing EI programmes which show effects on outcomes; (4) a randomised control trial will evaluate an emotional intelligence training intervention for social workers and assess its effect on social work performance and stress.

DATA DESCRIPTION (ABSTRACT)

Modes de partage des données de recherche

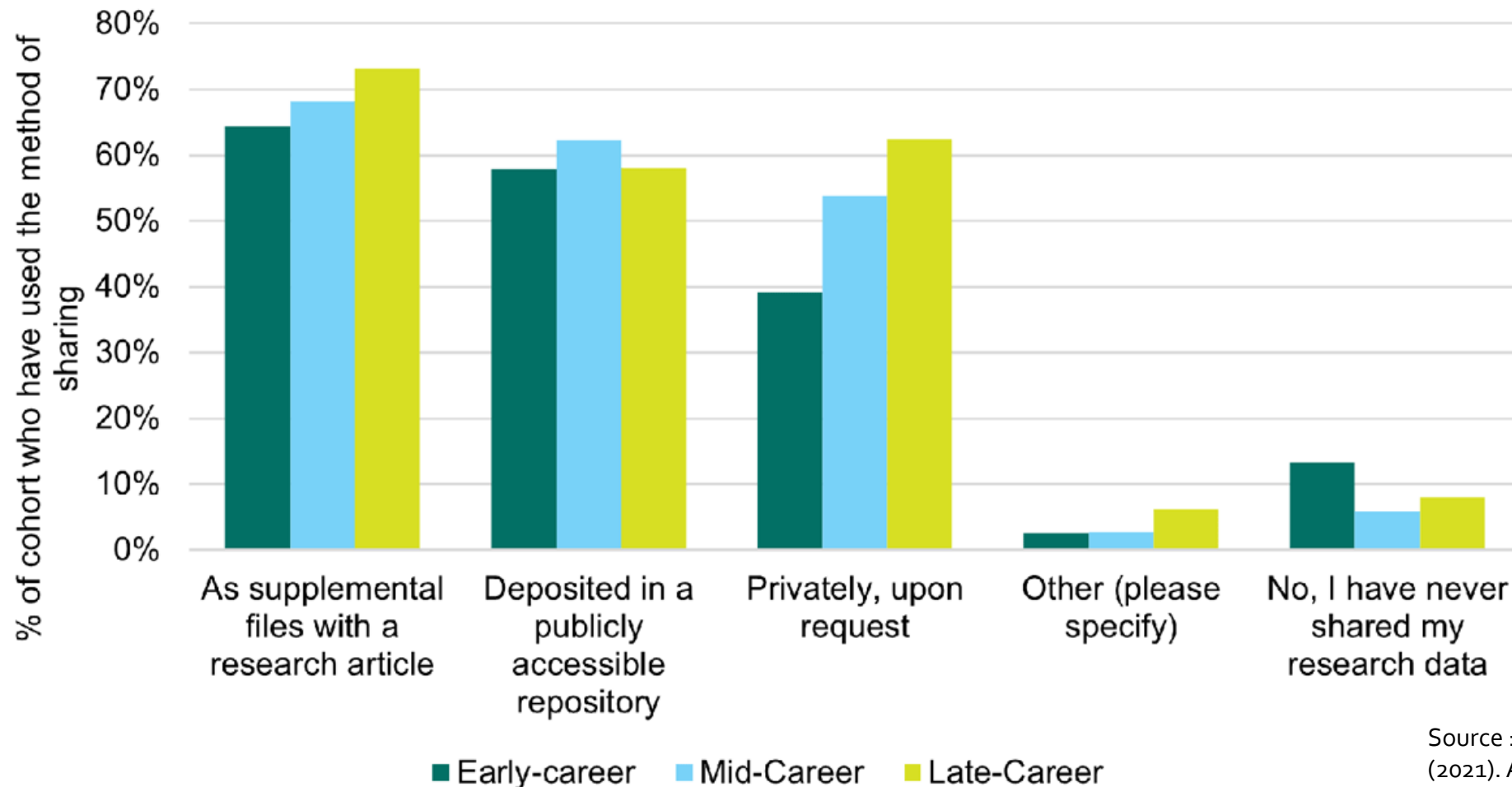
Quel mode de partage des données vous semble-t-il le plus fréquent?

- Fichiers supplémentaires à un article scientifique
- Dépôt dans un entrepôt de données accessible publiquement
- Privé, sur demande
- Autre
- Aucun partage

Quels avantages et quelles limites associez-vous à chacun d'entre eux?

Modes de partage des données de recherche

Methods of data sharing previously used



Source : Hrynaszkiewicz, I., Harney, J., & Cadwallader, L. (2021). A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Science Journal*, 20(1), 31. <https://doi.org/10.5334/dsj-2021-031>

Modes de partage des données de recherche

→ Enquête menée par PLOS

→ 1477 réponses

- › 20,9% Médecine et sciences de la santé
- › 30,4% Biologie et sciences de la vie
- › 9,7% Sciences sociales

“Sharing data as supplemental files alongside a research paper was the most common method for all career levels (67%), followed by deposition in a public repository (59%) and sharing privately on request (49%). Only 10% of respondents reported that they had never shared their research data – the largest proportion of whom (42%) work in Medicine and Health Science disciplines.”

Source : Hrynaskiewicz, I., Harney, J., & Cadwallader, L. (2021). A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Science Journal*, 20(1), 31.
<https://doi.org/10.5334/dsj-2021-031>

Modes de partage des données de recherche

NB ces modes de partage ne sont pas exclusifs les uns des autres, des fichiers supplémentaires à un article peuvent être déposés dans un entrepôt de données, parfois par la revue elle-même.

Exemple:

- Article : Rousselot, N., Tombrey, T., Zongo, D., Mouillet, E., Joseph, J.-P., Gay, B., & Salmi, L. R. (2018). Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners. *BMC Medical Education*, 18(1), 254. <https://doi.org/10.1186/s12909-018-1368-y>
- Fichiers déposés dans Figshare : Rousselot, N., Tombrey, T., Drissa Zongo, Mouillet, E., Jean-Philippe Joseph, Gay, B., & Salmi, L. (2018). *Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners* [Data set]. Figshare. <https://doi.org/10.6084/M9.FIGSHARE.C.4297076.V1>



Additional files

Additional file 1: of Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners

Additional file 1: Two parts of the tool to assess EBP skills: 1) Content of the skill assessment form; and 2) Scoring grid.

1) Skill assessment form

First part: Case vignette

Participants had 20 minutes to read a case vignette and answer four questions.

Summary of the case vignette:

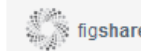
A 75-years-old man visits his general practitioner. In his medical history: an ischemic stroke 2 years before, atrial fibrillation, smoking, hypertension, and hypercholesterolemia. He was worried by a risk of epilepsy because of his stroke; asked if his use of coffee was excessive.

Additional File 1

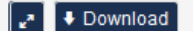
view download

Additional File 2

view download



1 / 2



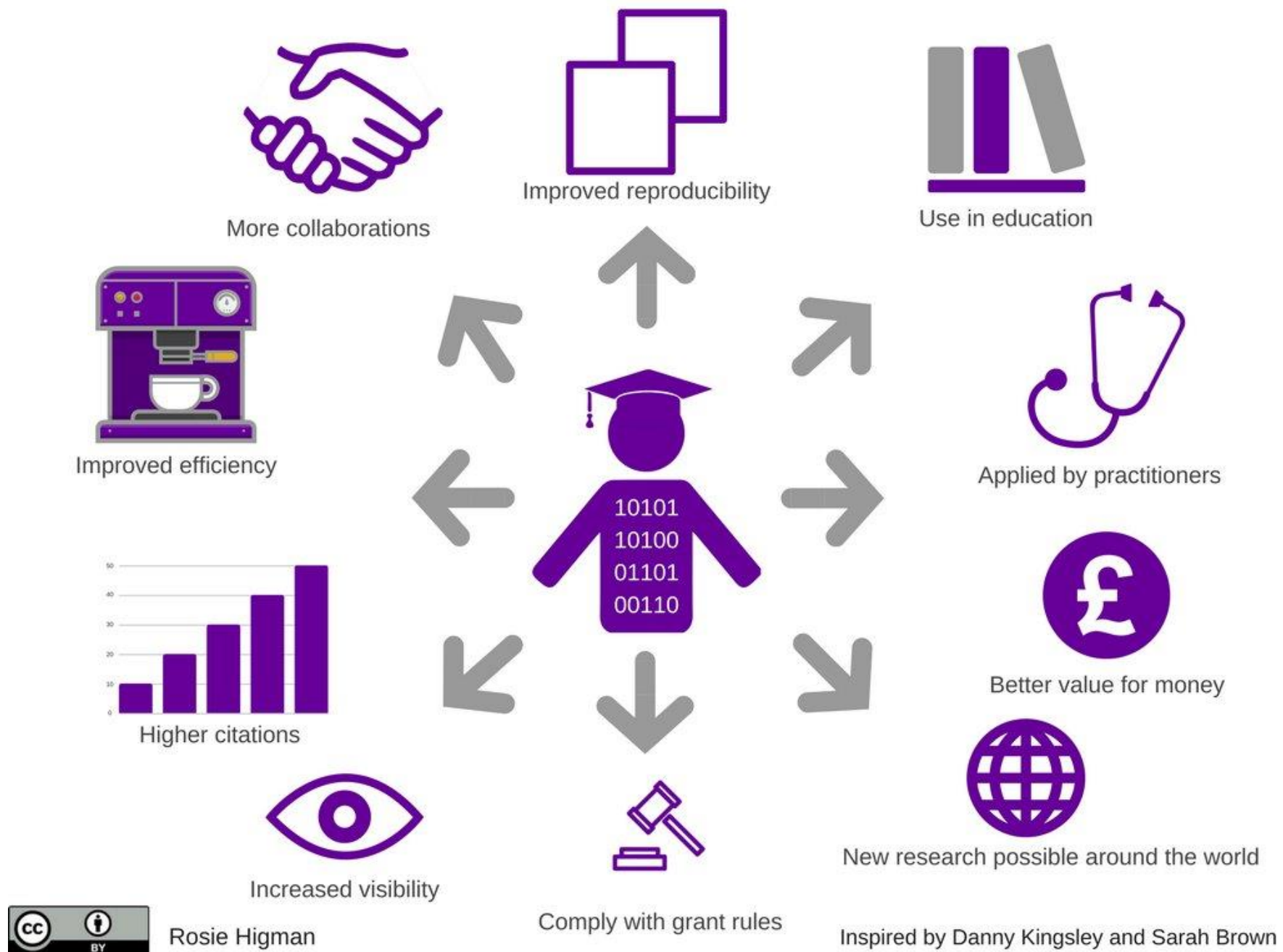
Additional File 1

Two parts of the tool to assess EBP skills: 1) Content of the skill assessment form; and 2) Scoring grid. This file gives more information about our tool. (DOCX 61 kb)




Additional file 1:

Two parts of the tool to assess EBP skills: 1) Content of the skill assessment form; and 2) Scoring grid. This file gives more information about our tool. (DOCX 61 kb)

Enjeux



Enjeux

Stage	Problem	Solution	Helps to know...	Agents
Data collection and analysis 	Waste of research (duplication)	Preregistration	Who is doing what, how and when?	Researchers to adopt it Journals to enforce it
	Ethical concerns	Preregistration	Is the study in line with good research practices?	
	Flawed studies	Registered reports	Is the study designed appropriately to answer the question of interest?	
Publication process 	Expedite reviewing	Open reviews	Are the findings verified by independent researchers?	Researchers to adopt it Journals to enforce it Institutions and funding agencies to value it Policymakers to allow it
	Distrust of published results	Data and code sharing	Can scientist obtain similar results from the same data?	
	Conflict of interests	Disclosure of conflict of interests, including editorial roles	Are the authors incentivized to publish these findings and to publish in this journal?	Researchers to adopt it Journals to enforce it
Communication 	Misuse of preprints Misleading headlines, exaggerations	Collaboration between journalists and scientists	Can the public trust scientific news in the media?	Journalists and news editors Institutions
	Paywalled manuscripts	Open Access on all manuscripts	The exact content of manuscripts used as a source. Can accelerate research through universal access to scientific findings	Researchers to favor it Policymakers and institutions to enforce it

Source : Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives : Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, 21(1), 117. <https://doi.org/10.1186/s12874-021-01304-y>

Enjeux - exemple politique d'éditeurs

“In addition to asking authors to link to shared data in the author note, many APA journals now require statements linking to shared data, materials, and/or code—or specifying their ethical or legal reasons for not sharing:

- Authors should provide a data availability statement indicating whether the data, methods used in the analysis, code, and materials used to conduct the research will be made available to any researcher for purposes of reproducing the results or replicating the procedure. In both the author note and at the end of the method section, either specify where that material will be available or note the ethical or legal reasons for not doing so. ”

Source : American Psychological Association. *Data sharing*. Consulté 29 novembre 2021, à l'adresse <https://www.apa.org/pubs/journals/resources/data-sharing#>

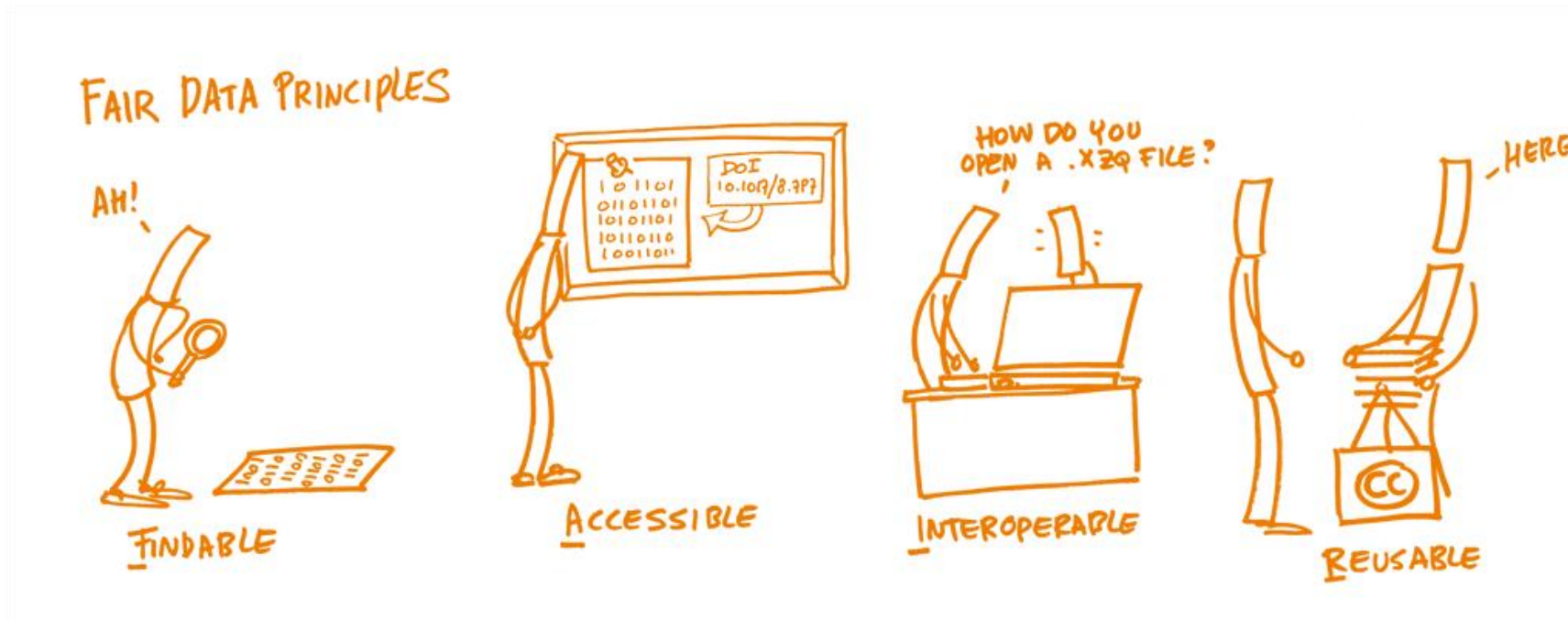
Entrepôts de données et principes FAIR

Principes FAIR -> les données doivent pouvoir être trouvées, comprises et réutilisées.

Cela ne signifie pas qu'elles doivent être nécessairement en accès entièrement libre.

Ces principes s'appliquent également aux métadonnées associées aux données.

Principes FAIR en détail -> [Page web « FAIR et DMP » sur le site de la BU](#)



Source image : [The Open Science Training Handbook](#)

Entrepôts de données et principes FAIR

DOI et citation :
Findable - Accessible

Documentation :
Reusable

Une documentation plus structurée et standardisée permettrait d'augmenter les scores Interoperable - Reusable

Signage interventions for stair climbing at work: more than 700,000 reasons for caution

Eves, Frank, University of Birmingham, <https://orcid.org/0000-0002-4921-7169>

evesff@bham.ac.uk

Publication date: October 25, 2019

Publisher: Dryad

<https://doi.org/10.5061/dryad.qz612jm91>

Citation

Eves, Frank (2019), Signage interventions for stair climbing at work: more than 700,000 reasons for caution, Dryad, Dataset, <https://doi.org/10.5061/dryad.qz612jm91>

Abstract

Increased stair climbing reduces cardiovascular disease risk. While signage interventions for workplace stair climbing offer a low-cost tool to improve population health, inconsistent effects of intervention occur. Pedestrian movement within the built environment has major effects on stair use, independent of any health initiative. This paper used pooled data from UK and Spanish workplaces to test the effects of signage interventions when pedestrian movement was controlled for in analyses. Automated counters measured stair and elevator usage at the ground floor throughout the working day. Signage interventions employed previously successful campaigns. In the UK, minute-by-minute stair/elevator choices measured effects of momentary pedestrian traffic at the choice-point (n = 426,605). In Spain, aggregated pedestrian traffic every 30 min measured effects for 'busyness' of the building (n = 293,300). Intervention effects on stair descent (3 of 4 analyses) were more frequent than effects on stair climbing, the behavior with proven health benefits (1 of 4 analyses). Any intervention effects were of small magnitude relative to the influence of pedestrian movement. Failure to control for pedestrian movement compromises any estimate for signage effectiveness. These pooled data provide limited evidence that signage interventions for stair climbing at work will enhance population health.

Methods

Observations from automated counters followed by bootstrapped regressions

Usage Notes

Read me file for UK interventions

Up from ground floor

The coding used to allow analysis of the file with logistic regression is explained below. The dependent variable is stair and escalator choice. If you have any queries, please do not hesitate to contact me directly (evesff@bham.ac.uk).


For ascent, file IJERPHMultiSiteUKUp, the coding is as follows below.

INTSTAGE; 1 =baseline, 2=first intervention, 3=second intervention.

BASE0INT1; the coding allows comparison of the baseline (0) with the first intervention (1).

PH2ASE1; the coding allows comparison of the first intervention (0) with the second (1).

Data Files

 Download dataset

October 25, 2019


IJERPHBarce...iSite.xlsx	276.45 kB
IJERPHMulti...KDown.xlsx	5.92 MB
IJERPHMulti...eUKUp.xlsx	6.48 MB


Related Works


Article

<https://doi.org/10.3390/ijerph16193782>

Metrics

 25 views

 0 downloads

 0 citations

Keywords

population health

signage effectiveness

stair climbing

workplace

License

This work is licensed under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license.



Licence
CC:
Reusable

Critères de choix d'un entrepôt de données

→ L'entrepôt est-il **certifié**?

- › En savoir plus sur la certification *CoreTrustSeal* : RDA France. (2019). Entrepôts de données de confiance : Critères de conformité. Repéré à <https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf>

→ Gère-t-il différents **types d'accès**? Permet-il par exemple un accès restreint ou sous **embargo**?

- › Y compris la gestion des demandes d'accès pour les données en accès restreint?
- › Y compris la possibilité de définir le type d'accès fichier par fichier? [cf. exemple ICPSR ci-dessous]

→ Prend-il en charge un **format de données** particulier?

→ Quelles sont les exigences ou possibilités en termes de **précision et structuration de la description des données déposées**? Devrez-vous ou pourrez-vous fournir des métadonnées riches?

→ Propose-t-il d'autres services particulièrement adaptés à mes besoins? [cf. exemple connexion Zenodo-Github ci-dessous]

ICPSR - gestion fine des accès

Outcome Evaluation of the Wisconsin Residential Substance Abuse Treatment (RSAT) Program: The Mental Illness Chemical Abuse (MICA) Program at Oshkosh Correctional Institution, 1997-2000 (ICPSR 3082)

Version Date: Mar 30, 2006 [Cite this study](#) | [Share this page](#)

Principal Investigator(s): [D. Paul Moberg](#)

[University of Wisconsin. Medical School. Department of Population Health Sciences. Center for Health Policy and Program Evaluation](#); [University of Wisconsin. Medical School. Department of Population Health Sciences. Center for Health Policy and Program Evaluation](#)

<https://doi.org/10.3886/ICPSR03082.v1>

Version V1

Download ▾

Analyze Online (0)

Access Restricted Data








At A Glance

Data & Documentation

Variables

Data-related Publications

Export Metadata

Name	Size	Preview	Download
DS0 Study-Level Files	4 MB		
DS1 Admissions Data	0 bytes		
DS2 Comparison Data	0 bytes		
DS3 Institutional Post-Graduation Data	0 bytes		
DS4 Termination Data	0 bytes		
DS5 Three-Month Post-Release Data	0 bytes		

Downloads *
[Usage Report](#)
* past three years

8
[Data-related Publications](#)

Notes

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.
- One or more files in this data collection have special restrictions. Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.

Exemple :
<https://doi.org/10.3886/ICPSR03082.v1>

Zenodo - connexion GitHub



Search

UploadCommunities

Log inSign up

June 8, 2018

SoftwareOpen Access

Analyses of Ebola virus disease outbreaks

Christian Althaus

This release contains data and  code files from the following analyses of Ebola virus disease (EVD) outbreaks:

- Althaus CL. (2014) [Estimating the reproduction number of Ebola virus \(EVOB\) during the 2014 outbreak in West Africa](#). PLoS Curr, 6.
- Althaus CL, Low N, Musa EO, Shuaib F, Gsteiger S. (2015) [Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control](#). Epidemics, 11:80-4.
- Althaus CL. (2015) [Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo](#). PeerJ, 3:e1418.
- Althaus CL. (2015) [Ebola superspreading](#)70135-0). Lancet Infect Dis, 15:507-8.
- Abbate JL, Murall CL, Richner H, Althaus CL. (2016) [Potential impact of sexual transmission on Ebola virus epidemiology: Sierra Leone as a case study](#). PLOS Negl Trop Dis, 10:e0004676.
- Althaus CL. (2018) [Real-time analysis of the 2018 Ebola outbreak in the Democratic Republic of Congo](#). GitHub.

Preview

Ebola-v1.0.0.zip

calthaus-Ebola-4888566

DRC (GitHub 2018)

Ebola_outbreak_DRC2018_data.csv1.0 kB

README.Rmd9.8 kB

README.md2.8 kB

README_cache

markdown_github

__packages33 Bytes

fit_ae1c24202d0029d622dc3ee7bdb65569.RData2.7 kB

fit_ae1c24202d0029d622dc3ee7bdb65569.rdb8.6 MB

fit_ae1c24202d0029d622dc3ee7bdb65569.rdx457 Bytes

reproduction_73d668b2a1a28489f76ef1d634598955.RData2.7 kB

reproduction_73d668b2a1a28489f76ef1d634598955.rdb42.9 MB

reproduction_73d668b2a1a28489f76ef1d634598955.rdx251 Bytes

69views

7downloads

See more details...

Available in



Indexed in



Publication date:

June 8, 2018

DOI:

DOI10.5281/zenodo.1285677

Related identifiers:

Supplement to

<https://github.com/calthaus/Ebola/tree/v1.0.0>

License (for files):

[Other \(Open\)](#)

Exemple :
<https://doi.org/10.5281/zenodo.1285677>



Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne UBx : « Trouver un entrepôt de données »

Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne UBx : [Trouver un entrepôt de données](#)

→ Pour identifier l'entrepôt de données généraliste qui répond le mieux aux besoins, à partir de la réponse aux 4 questions suivantes :

- › Allez-vous publier des données dont l'accès doit être restreint?
- › Recherchez-vous un entrepôt avec des options de dépôt gratuit?
- › Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données?
- › Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne?

→ Une brève fiche descriptive accompagne chacun des huit entrepôts de données comparés : [4TU.ResearchData](#), [Figshare](#), [Harvard Dataverse](#), [Dryad](#), [Mendeley Data](#), [Open Science Framework \(OSF\)](#), [B2SHARE](#), [Zenodo](#)

Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne : [Trouver un entrepôt de données](#)

Aide au choix d'un entrepôt

Trouver un entrepôt de données

Vous cherchez un entrepôt de données généraliste pour rendre accessibles vos données de recherche ?

Cet outil peut vous aider à trouver l'entrepôt qui répond le mieux à vos besoins. Pour identifier un entrepôt de données spécialisé pour votre domaine de recherche, interrogez l'annuaire **re3data** ou **contactez-nous**.

Pour plus d'informations concernant les données de recherche, consultez les pages de notre site web **Les données de recherche**.

Pour toute question, contactez-nous à l'adresse doc-soutienrecherche@u-bordeaux.fr

Cet outil est adapté du **Data Repository Finder** développé par l'université d'Utrecht. Il est en cours de paramétrage et de finalisation, les contenus et les fonctionnalités sont susceptibles d'évoluer.

- | | |
|---|--|
| 1. Allez-vous publier des données dont l'accès doit être restreint ?
<input type="radio"/> Oui <input type="radio"/> Pas nécessairement | 2. Recherchez-vous un entrepôt avec des options de dépôt gratuit ?
<input type="radio"/> Oui <input type="radio"/> Pas nécessairement |
| 3. Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données ?
<input type="radio"/> Oui <input type="radio"/> Pas nécessairement | 4. Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne ?
<input type="radio"/> Oui <input type="radio"/> Pas nécessairement |

Réinitialiser



Trouver un entrepôt de données spécialisé

Recommandations des éditeurs et des organismes de financement

Annuaire re3data

Service FAIRsharing

Recommandations

→ Editeurs : exemples [PLOS](#) et [Springer-Nature](#)

→ Organismes de financement

- › [ERC - European Research Council](#) : liste commentée d'entrepôts spécialisés par discipline - p. 12 et suiv.
- › [Commission européenne](#) : liste d'entrepôts généralistes et spécialisés par discipline - ces *Data Guidelines* comportent en outre des recommandations pour la préparation des données (notamment tabulaires).

→ Membres d'un consortium : CESSDA - Consortium of European Social Science Data Archives : <https://www.CESSDA.eu/About/Consortium>

Annuaire re3data

Registry of Research Data Repositories.

<https://www.re3data.org/>

Il s'agit de la ressource de référence, indexant plus de 2000 entrepôts. On peut rechercher par mot-clé et naviguer dans re3data par :

→ [sujet](#),

→ [type de contenu](#),

→ [pays](#).

The screenshot displays the re3data.org website interface. At the top, there is a navigation bar with links for Search, Browse, Suggest, Resources, and Contact, along with the DataCite logo. A search bar is located on the right side of the header. On the left side, there is a 'Filter' sidebar with various categories like Subjects, Content Types, Countries, AID systems, API, Certificates, Data access, Data access restrictions, Database access, Database access restrictions, Database licenses, Data licenses, Data upload, Data upload restrictions, Enhanced publication, Institution responsibility type, Institution type, Keywords, Metadata standards, PID systems, Provider types, Quality management, Repository languages, Software, Syndications, Repository types, and Versioning. The main content area shows search results for 'PubChem'. It includes a search bar, a pagination bar (1 to 108), and a 'Sort by' dropdown. The results for 'PubChem' are displayed with filters for Subject(s) (Basic Biological and Medical Research, Chemistry, Biology, Life Sciences, Natural Sciences), Content type(s) (Databases, Images, Structured graphics, Scientific and statistical data formats, Raw data, Plain text, Structured text), and Country (United States). A detailed description of PubChem is provided, mentioning its three databases: PubChem BioAssay, PubChem Compound, and PubChem Substance. Below this, the 'World Data System' is also listed with its own filters and description.

Annuaire re3data

De nombreux filtres permettent ensuite d'affiner les listes de résultats, en fonction par exemple des critères suivants.

→ **Certificates** : par quelle certification l'entrepôt est-il qualifié?

→ **Accès** : gradient de valeurs de fermé à ouvert - se décompose en 3 types d'accès :

- › **Database access** : accès à l'entrepôt de données lui-même : sous quelles conditions un utilisateur peut-il accéder à la base de données en général?
- › **Data access** : accès aux jeux de données déposés dans un entrepôt de données spécifique : sous quelles conditions un utilisateur peut-il accéder à un jeu de données?
- › **Data upload** : accès à la soumission de données : sous quelles conditions un utilisateur peut-il soumettre des données ?

→ **Versioning** : les jeux de données peuvent-ils être versionnés?

Service FAIRsharing

[FAIRsharing.org](https://fairsharing.org) répertorie non seulement des entrepôts mais également des standards, des méthodes, des vocabulaires, etc. Pour chaque ressource sont notamment spécifiés les critères suivants.

- › Le **statut** : en développement / opérationnel / incertain / déprécié
- › La **recommandation** : nom de l'éditeur, de la revue, etc. qui recommande

Voir : Sansone, S.-A. et al. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*, 37(4), 358-367. <https://doi.org/10.1038/s41587-019-0080-8>

The screenshot displays the FAIRsharing.org website interface. At the top, the logo 'FAIRsharing.org' is shown with the tagline 'standards, databases, policies'. Below the header, there is a red button labeled 'Recommended'. The main content area features a 'View as Table' and 'View as Grid' toggle, a 'Sort by' dropdown menu set to 'Best Match', and a section titled 'Recommended Records'. This section includes a red 'Recommended' header, followed by filters for 'Associated Publication?' (No Publication, Has Publication), 'Claimed?' (No Maintainer, Has Maintainer), and 'Record Status' (Uncertain, Deprecated, In development, Ready). At the bottom, there is a 'Standard Type' section with 'Reporting Guideline' (11) and 'Terminology Artifact' (7).



Merci pour votre attention

Des questions?

Contact : frederique.flamerie-de-lachapelle@u-bordeaux.fr