Data Sharing in Biomedical Sciences. A Systematic Review of Incentives

Running title: Data sharing and Biomedical Sciences

Thijs Devriendt¹, Mahsa Shabani¹², Pascal Borry¹

¹KU Leuven, Department of Public Health and Primary Care, Centre for Biomedical Ethics and Law
²Metamedica, Faculty of Law and Criminology, Ghent University, 9000-B Gent, Belgium

*Corresponding author: thijs.devriendt@kuleuven.be
Kapucijnenvoer 35 Box 7001
3000 Leuven
+32 16 19 37 25

Key words: science policy, ethics, data sharing, co-authorship, incentives
Abstract

**Background:** The lack of incentives has been described as the rate-limiting step for data-sharing. Currently, the evaluation of the scientific productivity by academic institutions and funders has been heavily reliant upon the number of the publications and citations, raising questions about the adequacy of such mechanisms to reward data generation and sharing. This paper provides a systematic review of the current and proposed incentive mechanisms for researchers in biomedical sciences and discusses their strengths and weaknesses.

**Methods:** PubMed, Web of Science and Google Scholar were queried for original research articles, editorials and opinion papers on incentives for data sharing. Articles were included if they discussed incentive mechanisms for data sharing, were applicable to biomedical sciences, and were written in English.

**Results:** Although co-authorship in return for the sharing of data is common, this might be incompatible with authorship guidelines and raise concerns over the ability of secondary analysts to contest the proposed research methods or conclusions that are drawn. Data publication, citation and alt-metrics have been proposed as alternative routes to credit data generators, which could address these disadvantages. Their primary downsides are that they are not well-established, it is difficult to acquire evidence to support their implementation and that they could be gamed or give rise to novel forms of research misconduct.

**Conclusions:** Alternative recognition mechanisms need to be more commonly used to generate evidence on their power to stimulate data sharing, and to assess where they fall short. There is ample discussion in policy documents on alternative crediting systems to work towards Open Science, which indicates that there is an interest in working out more elaborate meta-science programs.
Background

The Open Data movement aims to promote data sharing and to facilitate free and equitable access to research data underlying publications and studies (1). The sharing of research data within biomedical research is envisioned to bring considerable advantages, such as allowing re-analysis of data, testing secondary hypothesis, performing meta- or co-analyses of multiple data sets, or using data for educational purposes (2). Furthermore, data sharing allows avoiding unnecessary duplication of studies and brings “benefits of scale”, such as quicker discovery processes and higher statistical power. It would therefore accelerate translational research and inform better the development of policies and regulations in, for example, the public health setting (2–5). In addition, data sharing can increase the transparency of the research process, increase the accountability of researchers and incentivize researchers to engage in better data documentation and research methods (2). Augmented data usage also means that the contributions of research participants are more strongly recognized and appreciated (6,7). Routine data sharing could bring substantial benefits to the field of biomedical sciences as it reduces waste in research and improves the return on investment for society.

In recent years, numerous high-profile organizations have underlined the need for more data sharing of publicly funded research, including the Organization for Economic Co-operation and Development (OECD), the European Commission and the National Institutes of Health (NIH) (8–10). In 2015, the European Union launched the Open Research Data Pilot in the Horizon 2020 (H2020) program in several areas and subsequently extended the pilot to cover all areas of the program in 2017 (11). However, researchers were initially reluctant to engage in the pilot program and often decided to opt-out, as there were no incentives to dedicate time and funds to data preparation and sharing (12). Many facets of the policy challenges for Open Science are extensively described in Open Science by Design, issued by the National Academies of Science, Engineering, and Mathematics (13). It addresses barriers and facilitators for Open
Science, how policies and practices of actors in the research enterprise are influencing Open Science practices, and touches upon solutions within several policy areas, such as incentives and infrastructure. The movement towards Open Data advances with incremental steps and at differential speeds within scientific disciplines, owing to the fact that different types of data and academic culture pose unique challenges for data sharing within different fields. Thus, no “one-size-fits-all” solution exists across all scientific disciplines (14).

Despite Open Science ideals being increasingly in the spotlight in recent years, data sharing has not yet become a common practice within the biomedical sciences. One report on data sharing during public health emergencies (e.g. pandemics) showed that only 31% of studies provided access to all data underlying the paper within the article itself, with another 5% stating that data would be provided upon request (15). The report concluded that, even during the 2014 Ebola crisis in West Africa, data sharing stayed the exception and not the norm. A review of 318 biomedical journals by Vasilevsky et al. concluded that only 11.9% stated specifically that data sharing was required as a condition for publication (16). Furthermore, it has been reported that many published papers do not comply with the data availability instructions of the journals they were published in; for example, the data sharing rate of all primary raw data for clinical trials was 24% for those articles subject to BMJ’s data sharing policy (17).

Within the academic literature, many barriers have been recorded why researchers may not share data. One such barrier is the fear of getting “scooped”, which means that results of competing analyses are published before those that have produced the primary data (hereafter “data generators”) have published theirs. Other concerns include potential misinterpretation of data by other scientists, the misuse of data by commercial companies, the lack of financial resources or specialized personnel for sharing, privacy concerns, or restricting legislation (14,18,19). Some fields are also lacking infrastructures that would allow data sharing while adhering to the FAIR (Findable, Accessible, Interoperable, Reusable) principles (20).
Furthermore, the lack of professional rewards might discourage scientists from sharing data, which has been described as the “rate-limiting step” for data sharing (21,22). Fecher et al. identified the drivers of academic data sharing through a systematic review of the literature and a survey, and outline their view that academia should reinforce incentives for data sharing through formal recognition and potentially financial reimbursement (23). In this context, alternative recognition systems have been regularly proposed as solutions. Nevertheless, Rowhani et al. showed through a systematic review that empirical evidence underlying these incentives is scarce, as only the usage of data sharing badges can be supported by evidence from a pre-post design (24).

In light of the lack of incentives for sharing data which has been reported in the literature, we set forth to systematically review which incentives exist and have been proposed, and to critically assess their advantages and disadvantages. We conceptualize incentives as recognition mechanisms that can incentivize researchers to share data.

**Methods**

**Data sources**

The databases of PubMed (MEDLINE) and Web of Science were queried for articles. Additionally, the first 30 pages of Google Scholar (300 records) were explored to complement the search. This cut-off was manually chosen due to the number of records identified by Google Scholar (>15.000). The following search string was employed for the queries: ('data sharing" OR "datasharing" OR "Open Science") AND (incentive OR incentives). The search was carried out between May and June 2019. No limits on publication dates were used. The reference lists of selected articles were consulted to retrieve and include additional relevant articles.

**Study selection**

Articles were included in the systematic review if they complied with the following criteria: (1) the topic must be (at least partially) on incentives for data sharing in research; (2) must be
applicable to biomedical research; (3) articles in peer-reviewed journals in English. Articles were excluded if they addressed solely other branches of science (e.g. physics), addressed data sharing between institutions in public health care systems (e.g. HMOs) or were classified as reports/policy documents. After selection based on title and abstract screenings, two persons (TD and MS) independently evaluated the full texts in order to determine their suitability for inclusion. In cases of disagreement, articles were revisited and differences in interpretation discussed to reach consensus. Our systematic review identified 20 articles that complied with the inclusion criteria of the study (Table 1). The PRISMA flowchart details the study selection process (Fig 1).

Data extraction and synthesis

Quotes were extracted from the text and coded, and the extracted codes have been discussed and verified by TD and MS. The results were extracted using an inductive approach. Results were classified in thematic categories, such as incentives for data sharing and then conditions for implementation of these incentives. The structure of the categories was developed independently by two authors (TD and MS) and discussed, and subsequently a single thematic structure was developed.

Results

We report our results into various categories of incentives that were identified in the literature: (a) co-authorship; (b) embargos; (c) data journals and data publications; (d) alt-metrics; (e) data sharing badges; and (f) data advertisements.

Co-authorship

Many of the articles included in this review address the situation where data generators are included as co-authors in return for sharing their research data with secondary users. Co-authorship requests have traditionally been a standard practice to credit researchers for data generation and sharing, and could be considered a major incentive to share data.
A number of papers reported various practices regarding co-authorship in the context of data sharing consortia among data producers (25–28). Some, such as the Alzheimer’s Disease Neuroimaging Initiative (ANDI) require co-authorship for any manuscript generated from their data while others, such as the 1000 Functional Connectome Project (FCP), require data generators only to be mentioned in the acknowledgements section (28). Multiple papers address the fact that co-analysis of data of different sources is becoming increasingly frequent and that this has led to expanding author lists for data sharing consortia (25,28).

Co-authorship has been evaluated differently across the included papers. Some papers supported co-authorship arguing that requesting authorship in return of data sharing is justifiable or “an incentive and an ideal” (28,29). Ohmann et al. underline the advantages that actively collaborating with data generators might bring, namely that they can share key insights with the secondary users about the study design, the data or the analyses (30). Lo et al. share this view by stating that “secondary users can better understand the data set if they work closely with trialists who planned the trial and managed the data”, adding that “the prospect of co-authorship also gives trialists an incentive to document the data set in sufficient detail to assist with secondary analyses that might be done years later” (29).

Others look upon these practices less favorably, calling it abusive and solely in use because of the lack of other established means to credit data generators (11). Various articles refer to the downsides and extra complexities that co-authorship in return for data sharing brings (3,25,28). For example, there are no guidelines on how many co-authors are justifiable, on how many publications these co-authors should be included and what to do when author lists become excessive (25,28). Gorgolewski et al. indicated that it might be inappropriate to include individuals that have solely contributed to data generation as co-authors (28). Another concern that is being raised by the authors is that data users and generators may also disagree fundamentally on the method of analysis or interpretation of results, and there is no way to
easily resolve these issues (28). Therefore, Bierer et al. have argued that co-authorship might discourage interpretations deviant from the original papers created by the data generators. They also mention the need for collaboration with primary data generators as cumbersome or infeasible in practice (3).

Some held more pragmatic views on co-authorship, stating that, although there is no necessity to involve data generators, there may be benefits to involve the data generators in the analysis process or that co-authorship is acceptable under certain circumstances (30,31). Factors that are argued should be taken into account include: (i) the degree to which the data are integral to the manuscript; (ii) the unique or novel nature of the data; and (iii) the willingness of the data generator to participate in the drafting and revision of the manuscript and to take responsibility for the final product (31).

**Embargos**

An embargo (or moratorium) in this context means holding a monopoly on publications from a certain data set within a fixed time period. As pointed out by Bezuidenhout et al., embargos are a way to deal with researchers’ concerns that rapid data sharing might lead to other groups publishing important research results before them (14). As such, embargos are limited in both time and scope, and essentially safeguard the “right of first publishing” of the data generators.

Several articles refer to the disadvantages of implementing embargos. Dyke et al. cite that one danger of embargos is “unintentionally delaying analyses by other groups” (5). Another concern is raised by Kaye et al, which is that embargos can also put immense pressure on research groups, which should analyze the data set, publish (and in some cases, generate additional data) within short timescales (25).

The appropriate length of an embargo might be problematic to determine in some cases. Ohmann et al, state that it is hard to define an exact timeline for disclosure of patient data from
clinical trials. Among other issues, the release of data should be dependent upon the possibility and timing of primary publications, the complexity of the study, the nature of the data, the amount of preparation the data might require, and the access regime under which it is planned to be made available (30). Furthermore, there may be concerns of researchers working in low-middle income countries (LMICs) over data sharing. As researchers in LMICs generally need longer times to generate, analyze and make a publication out of the data, short embargo timetables could still lead to concerns over researchers of high-income countries scooping their primary findings (14).

Data journals and data publications

Recently, publishing data sets as citable entities that detail the experimental protocol and data specification (i.e. metadata) has become possible with the emergence of data journals. Data papers require that data be deposited into a managed repository and that a traceable identifier such as a digital object identifier (DOI) is assigned to data, and allow researchers to directly obtain credit for generating their data set (i.e. through an additional publication or citation). Data citation can be achieved by submitting to repositories which automatically attach data DOIs.

Choudhury et al. point out that the publishing of data sets in data journals is common in genetics, robotics and environmental sciences (32). Nevertheless, Gorgolewski et al. state that infrastructure (e.g. specialized venues) for routine publication and dissemination of imaging data sets is still lacking (28). Additionally, Dallmeier et al state that scientific communities may not yet recognize published data as a citable publication (18).

The move towards data publications is considered to be a positive evolution by several authors, owing to the fact that it allows researchers to have an additional publication and receive downstream citations for re-used data sets (33–35). To illustrate, Kattge et al. describe data publishing as the realization of “true benefits for researchers in return for the time-consuming
task of making their data and metadata widely usable” (35). Furthermore, they also underline some additional benefits may accompany publishing of data sets, such as incentivizing the collection of high-quality data by researchers (35). With regard to citation metrics, Choudhury et al. state that the h-indices might already be suitable to be used to measure the results of data sharing, as long as data are routinely cited from repositories for secondary or re-analyses by data users other than the original contributors (32). Data publication and subsequent citation therefore constitutes one of the pre-eminent ways to acquire credit for data re-use.

Nonetheless, the introduction of data papers may bring its own limitations and concerns over misuse. One concern with data publication raised by Gorgolewski et al. is that it might lead to slicing up the data set into separate subsets, which would then each be accompanied with a data publication. Then, the interpretation of the data might be disturbed as only a view on the entire data set can clarify the rationale of the intended experimental design and methodological choices (28). Ohmann et al. bring forth another concern, which is that different versions of data sets and documents may be circulating, such as with long-running studies where the protocol might have been updated (30).

**Alt-metrics: BRIF, the S-index, data authorship**

Scientific output has been traditionally quantified in the academic system through the number of publications and citations (36). Ferguson et al. argue that as long as data are still regarded as supplements to the scientific record rather than primary products of research to be credited, cited and tracked, researchers will weigh the time and resources required for preparing data for release to the (limited) personal benefits they are likely to bring (34). To resolve this issue, some have called for the introduction of ‘alt-metrics’ that more broadly and accurately attempt to capture scientific output. With regard to data sharing, several such alternative metrics have been proposed in the literature.
Ascoli et al. suggest that the S-index could incentivize researchers to share data and code and can be used independently from the existing publication indices (33). Olfson et al. conceptualize the S-index to be quasi-analogous to the h-index. They describe the S-index as follows: “For each investigator sharing data, publications using their shared data would be ranked in descending order by number of citations and the value of their S-index would be the number of papers (N) in this list with N or more citations” (6). Thus, all papers using data disseminated by the author are ranked by number of citations, and N is determined by seeking the maximum number of papers (N) with at least N citations. The authors point out that the S-index would be compatible with the recent proposal to acknowledge “data authors” separately from regular authorship (see below) (6). Kattge et al., however, hold the view that no distinction should be made between data and publication citations when developing performance metrics. They argue that creating a distinction between the two, such as through a “data h-index”, carries the risk of data contribution becoming a second-class performance measure (35).

Another proposal is the inclusion of the position of data author in publications separate from traditional authorship. Kaye et al. propose the idea conceptually by describing possible solutions for papers with extensive author lists: “An alternative means might be to make a distinction between a contributor, who has provided the data set, and an author, who has worked on the analysis or result.” (25). Bierer et al. describe the designation of data authorship in detail and state that data authors are responsible for the integrity of the data set but not for the results of the analysis (nor the conclusions) contained within the manuscript. When data generators actively work together with secondary analysts, the data generators can be designated as both regular and data authors (3).

Mabile et al. propose the BioResource Impact Factor (BRIF), which can be used to more accurately measure the different scientific value that different data sets (or more broadly, bioresources) have. The rationale is that metric tools for data sharing often solely rely on citations of data, and that this does not accurately reflect the scientific value (or utility) as other
parameters need to be taken into the equation. The authors give the example of bioresources of rare diseases, which only a smaller research community will utilize. In that case, the number of derivative publications and citations will be naturally low in comparison to more popular fields yet such research work may have significant clinical impact. The combined parameters then are translated into an impact factor for the bioresource, which can be used to capture the "success" of the enterprise (11).

Data sharing badges (journals)

Rowhani-Farid et al. elaborate on the handing out of data sharing badges by journals. The journal Biostatistics introduced “kitemarks” for data sharing in 2009, which meant that articles for which data or code was available were rewarded with the letter D (for Data), C (for Code) or R (for Reproducibility) on the front page of the PDF version of the article. For the journal Biostatistics, this resulted in a 7.6% increase in the rate of data sharing of health data within repositories (37).

The authors state that the main advantage of data sharing badges is that they are low-cost. The main disadvantage is that there may be misallocation of badges, which implies that it may not always be clear under which conditions allocation should be carried out. Verifying strict conditions for allocation however may be associated with an additional administrative burden (37).

Data advertisement (consortia)

Within consortia, Pisani et al. describe that it is at times arduous to draw researchers over the line to share their data. As such, coordinators of consortia have considered how providing incentives might facilitate data sharing between members. Toga et al. describe how the Global Alzheimer’s Association Interactive Network (GAAIN) provides incentives to participants to join their data sharing network. One way GAAIN has used to promote data contribution is to display the identities of all its data partners, including their logos and URL links that forward
investigators to their web sites, on each GAAIN search page (26). The conceptual premise of this “data advertisement” is that central listings of data enhance the use of data sources through increased visibility. Thus, this will lead to an increase in the use of data sets (and subsequent citations or collaboration opportunities) as an incentive (5).

**Discussion**

This systematic review suggests that to date, several crediting mechanisms have been proposed or implemented, each with their own strengths and weaknesses. While mechanisms such as co-authorship, data citation, and alt-metrics aim to offer direct credit to data producers, others such as data advertisement and data sharing badges aim to bring more visibility to data sets and reputational credit.

Crediting the data generator via co-authorship is a common practice to encourage fellow researchers to share data. Similar to this *quid pro quo* taking place in the sharing of data, the phenomenon also exists in the biobanking community where samples are being “traded” for co-authorship (38). This practice relies on recognized crediting mechanisms in academia and researchers remaining in control over “their” data or samples. Some authors prefer collaboration and shared authorship on resultant publications (39,40). The primary reasons for this are that without such support, valid analysis and data interpretation could be difficult when dealing with complex data, and that it would be unfair towards the data generators not to include them in such work (39,40). The International Charter for principles for sharing bio-specimens and data states that for hypothesis-generated data, researchers should be given time to explore and confirm initial findings, and that an offer of authorship could be included if a significant contribution to the acquisition, analysis or interpretation of data is made (41). Within the context of clinical trials, some have stated that collaboration with data generators is preferred yet where this is not possible, practical or desired, other crediting mechanisms need to be developed and implemented (42).
Providing authorship to all data contributors could easily lead to conflicts with authorship standards (43,44). Researchers might expect authorship every time (a subset of) their data is being used. Nevertheless, some have argued that this is unreasonable, especially in those cases where many subsets are used for pooled analysis, as this would lead to publications with extensive author lists (45). The significant growth of articles with such author lists, sometimes referred to as hyper-authorship, has been attributed to increased data sharing within biomedical sciences (46). Analyses of large pooled data sets can also be complex and very time-consuming, and the data generators may have limited understanding of the methodology or interpretation of results of the analysis (45).

Data publication, citation and data authorship could in principle overcome the hurdles of granting co-authorship to data contributors. Nevertheless, there is presently little empirical evidence that support alternative crediting schemes and they are primarily being discussed hypothetically (24). To procure evidence of their influence on data sharing practices, they would need to become further integrated into policy processes. However, policy makers may consider that the presence of empirical evidence needs to precede large-scale implementation, essentially creating a “catch-22 situation”.

Currently, data citation practices are not employed widely in the biomedical field. An increasing number of general-purpose data repositories such as Zenodo and Figshare now provide DOIs to stored data. However, one recent study showed that informal data citation is still much more common than formal data citation which include references to data DOIs (47). This means that citation practices still need to change in order to properly credit data generators and take advantage of data DOIs. The organization DataCite has initiated the aggregation of data citations and data usage metrics for data sets uploaded in indexed repositories (48). The publishing of data papers has also become more common, with one study documenting the existence of over 100 data journals in 2015 of which the majority were situated within medicine (49).
On another level, the specification of author roles has been suggested. Data authorship has not been formally recognized by funders/institutions, yet some authors have recently called for the integration of this position into hiring, promotion and tenure decisions (22). However, the CRediT taxonomy, which offers 14 detailed contributorship roles, has been adopted by many biomedical journals (50). The purposes of the CRediT taxonomy are manifold, include supporting adherence to authorship policies, clarifying responsibility for scientific products and enabling the development of new indicators of research value, use and re-use, credit and attribution. These two concepts offer different levels of granularity and could both support future role-dependent evaluation of scientists. However, the CRediT taxonomy might be less suitable to link data generators with datasets in the context of data sharing. This is because datasets do not possess IDs that are consistently linked with researcher IDs throughout subsequent reuse of the dataset (40).

Data citations, data papers and specification of authorship roles could give rise to novel metrics. Nevertheless, one criticism to further “quantification” through metrics is that it might lead to questionable research practices such as “data salami-slicing” (i.e. where authors might cut up datasets in subsections to publish them separately), “data self-citation”, “honorary and ghost data-authorship” or other forms of misconduct (51). Thus, a careful assessment of the potential adverse implications regarding research integrity and gaming is necessary before implementation (51).

The likelihood of these incentives being adopted will differ within subfields of biomedical sciences. There are several reasons that can explain this. First, privacy-sensitive patient data and non-privacy sensitive biological data differ as to where they might be uploaded. The former can be subject to closed access models (i.e. upon request) while the latter could be available in general repositories. Thus, it would be somewhat more difficult to collect data citations or usage metrics for the former. Second, the effort in terms of time and resources expended on data generation varies substantially between data types. This has as a result that stronger incentives might be necessary to compensate larger efforts. Third, if secondary analysis of
data requires active and regular input from the data generators, traditional incentives such as co-authorship could be more appropriate.

The problems around credit and incentives for Open Science practices within the academic system are increasingly being discussed on the policy level. For example, one of the tasks of the Working Group on Rewards under Open Science of the European Union has been to “propose new ways/standards of evaluating research proposals and research outcomes taking into consideration all OS activities of researchers, possibly recommending to pilot them under certain calls of Horizon 2020”. Their proposed Open Science Career Assessment Matrix (OS-CAM) includes the evaluation of research outputs by (1) using the FAIR principles; (2) adopting quality standards in open data management and open data sets and; (3) making use of open data from other researchers. They encourage and recommend the initiation of feasibility and pilot studies implementing OS-CAM to see whether the projects achieve proper quality of science, research integrity and greater public and peer engagement (12). The European Commission Expert Group on Altmetrics recommends that existing metrics (for data sharing, but also for regular publications) need to be used better (Recommendation #3), and that before introducing new metrics into evaluation procedures, the benefits and consequences need to be studied as part of a “meta-science” program (Recommendation #9) (52). The Open Science by Design report also proposes that research institutions should establish a culture that better rewards and supports researchers engaged in open science practices, and that funders should provide support for practices and approaches that facilitate this shift in culture and incentives (13).

With these policy evolutions in mind, alternative crediting mechanisms in support of Open Science should be explored further. Empirical evidence should be gathered on their capacity to steer data sharing behavior. The shortcomings of these systems should also be recognized. Aside from lack of incentives, other barriers for data sharing, such as lack of infrastructure or resources, should concurrently be identified and addressed within specific contexts. This is
equivalent to tackling problems structurally, rather than attempting to hard-force data sharing solely through policy actions and regulations. Over time, the primacy of authored articles as the sole indicator of productivity should then be reconsidered in favor of more balanced models that integrate other outputs such as code or data.

Conclusion
Co-authorship, embargo’s, data journals and data publications, alt-metrics, data sharing badges and advertisements were identified as incentives for sharing data within this systematic review. Co-authorship has drawbacks related to research integrity, which role specification (CRediT or data authorship) or data publication and citation could potentially resolve. However, they would need to become more standard practice as to generate evidence on their ability to stimulate data sharing. Moreover, the potential pitfalls of these mechanisms first need to be assessed. The discussions on alternative crediting systems to stimulate Open Science goals that are taking place at the policy level indicate that that there is an interest in working out meta-science programs. Novel incentives should be developed at the same time as aiming to address other barriers for data sharing (e.g. lack of resources/infrastructures). If such methods fall short to stimulate data sharing, stronger policy mandates and regulation might be necessary.

Limitations
This systematic review has several limitations. First, the review only addresses the academic literature while some incentives have been discussed more thoroughly in policy documents (e.g. data citation). Second, the literature on data sharing is very heterogeneous in general and is addressed in the fields and subfields of information sciences, (bio)ethics, bioinformatics, sociology, life sciences and medicine. Each field brings unique perspectives on the topic and the prevalence of each “publication type”, such as conceptual or opinion pieces, differs considerably. This complicates conducting a systematic review because of differences in
terminology (incentive vs. inducement) and commonly used databases. Third, understanding the manner in which incentives affect behavior is not straightforward and requires a deeper understanding of the conditions that restrict sharing within specific subfields. This makes summarizing the advantages and disadvantages of incentives/recognition systems harder. Lastly, many articles screened addressed incentives for data sharing only marginally and did not propose concrete solutions aside from referencing some that have already been proposed. The inclusion criterium that the article should address incentives at least in part therefore often raised discussions on when articles were actually providing sufficient information to be included. We consider that these are important methodological limitations.

Abbreviations

ANDI: Alzheimer Disease Neuroimaging Initiative  
BRIF: BioResource Impact Factor  
DOI: Digital object identifier  
FAIR: Findable, Accessible, Interoperable, Reusable  
FCP: 1000 Functional Connectome Project  
GAAIN: Alzheimer Association Interactive Network  
HIC: High-income countries  
HMO: Health Maintenance Organization  
LMIC: Low-middle income countries  
NIH: National Institutes of Health  
OECD: Organization for Economic Co-operation and Development  
OS: Open Science  
OS-CAM: Open Science Career Assessment Matrix

Funding

This publication is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825903 (https://ec.europa.eu/programmes/horizon2020/en). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Acknowledgements

TD, MS and PB conceptualized the research project. TD and MS carried out the empirical work. MS and PB supervised the project. TD wrote the draft. MS and PB edited and revised the draft. MS and PB acquired the funding for carrying out this project. All authors read and approved the final manuscript. The authors declare no existing conflict of interest.

References


15. Pisani E, Ghatuare A, Merson L. Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats. 2018. doi:10.6084/m9.figshare.5897608.v1


19. Ferguson L. Researcher Data Sharing Insights. 2020. Available from: https://content.cdntwrk.com/files/aHVzPTgzODM3JmItZD1pdGVtZWRpdG9yaW1hZ2UmZmlsZW5hbWU9aXRlbWVkaXRvcmllTWdIYXJhNjdINjc4YzE2MzQuanBnJnZlcNp b249MDAwMCZzaWc9YzNhZjhiNjdINjdIYzY2Q3MmQ0YjIzZjdlOGUwZWI2Zjg%253D


