Exploratory analysis of policy document sources in Altmetric.com and Overton

Biegzat Murat*, Ed Noyons** and Rodrigo Costas***

*b.mulati@cwts.leidenuniv.nl
0000-0001-7828-2472;
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands

** noyons@cwts.leidenuniv.nl
0000-0002-3103-2991;
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands

*** rcostas@cwts.leidenuniv.nl
0000-0002-7465-6462
Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
DSI-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP),
Stellenbosch University, Stellnbosch, South Africa

Abstract

Policy documents are one of the altmetric sources most crucial for comprehending the interaction between science and policy. In policy documents, the source is usually the institution or organization that published the given policy document. In this study we compare the policy document sources indexed by Altmetric.com and Overton. Altmetric.com is an altmetric aggregator that has been around for over a decade. Overton is a newer database aiming at being the most complete collection of international policy documents. Our findings reveal that Overton covered more policy organizations than Altmetric.com, although the overlap in sources is quite small. This low overlap may suggest that both data aggregators may be using slightly different operationalizations of their notion of "policy documents", which calls for more transparency on what notion of policy documents and policy organizations are considered by them. Regarding policy organizations indexed by both data aggregators, most policy documents are published by government and non-profit organizations. Future research should delve more into the different policy organizations that are being covered in Overton and Altmetric.com, as well as their geographical distribution.

Keywords: Altmetrics; Policy Document; Policy Document Source; Policy Organization

1. Introduction

With the development of digital scientific communication, new types of scholarly outputs have been developed and novel ways of academic communication have emerged. Different methods of studying the social impact of scholarly papers have become more prominent and diverse in the scientometric landscape. One important way of monitoring the societal relevance of research is policy document mentions to scholarly papers. Policy document mentions may shed light on the relationship between academic research and policy making (Bornmann, Haunschild & Marx, 2016).

In the past few years, different commercial initiatives have started to collect, index and aggregate policy documents and their citations. One of the early developments was Altmetric.com (Fang et al, 2020a), and more recently Overton (Szommszor & Adie, 2022) claims to be the world's largest collection of policy documents [1], being one of the most prominent databases for policy document citation analysis (Fang et al, 2020b). The existence

---

[1] https://help.overton.io/article/what-is-overton/

of these two concurrent data aggregators, seemingly indexing the same type of *event* (i.e., policy document citations) raises questions both conceptual (what definition and notion of "policy document" are they considering? Do both of them operationalize policy documents in the same manner?), as well as technical (how are policy documents identified and processed? what policy sources are being – or not – covered by each of the databases?). To the best of our knowledge such questions have not yet been discussed in the literature. Having a good understanding of what is being considered as policy document and policy source by these aggregators, and what do they tell us about the interactions of science and policy, are critical elements in the quest of reliable and valid indicators of the policy impacts of science.

In this study we aim specifically at analysing the coverage of policy sources included in these two aggregators, Overton and Altmetric.com. We frame our analysis in the context of the definition the role of a policy document in the context of exploring the interaction between science and policy.

1.1. A brief overview of the Altmetric.com

Altmetric.com was founded in 2011 by Euan Adie, for a simple purpose: to help researchers see the influence of their work in real time (Konkiel, 2020). In its first ten years, Altmetric.com has already been studied in in different ways, such as their correlation with citations (Costas, Zahedi, & Wouters, 2015), its quality, and developing also some specific indicators, like Twitter indicators (Na, 2015), Facebook indicators (Syn & Oh, 2015) or policy document indicators (Bornmann, Haunschild, & Marx., 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018).

1.2. A brief overview of Overton

Overton, founded in 2019 also by Euan Adie, is a novel database established with an explicit goal to increase the coverage and comprehensiveness of policy-focused altmetrics. Overton indexes policy documents from all over the world, and it is built by combining a broad panel of government sources with web crawling. Overton indexes legislative and executive documents, including governmental white papers, transcripts of parliamentary sessions, publications by think tanks, etc. Close to 75% of these records are provided by US, UK, and intergovernmental sources (Pinheiro, Vignola-Gagné, & Campbel, 2021). It is created by the web crawling of publicly accessible documents published by a curated list of over 30,000 organizations including governments, intergovernmental organizations, think tanks, and charities (Szomszor & Adie, 2022).

1.3. On the difference between Altmetric.com & Overton

Although Altmetric.com and Overton were founded by the same person, there seem to be some differences between Altmetric.com and Overton in terms of what they consider as policy documents and the policy sources they are indexing.

The definition of policy document

To consider whether policy documents are an appropriate source to study the role of research outputs to policy making, we need to know more about what is meant with the notion of "policy documents". Since policy documents can be quite different in nature, and very likely they may provide different information on the interactions between science and policy, the critical question then becomes *how are policy documents defined by Altmetric.com and Overton?* Interestingly, both data aggregators provide only rather loose definitions of what they mean with policy documents. In the case of Altmetric.com they claim to curate "policy sources that

are designed to change or otherwise influence guidelines, policy or practice"[2], tracking documents ranging "from government guidelines, reports or white papers; independent policy institute publications; advisory committees on specific topics; research institutes; and international development organizations". In the case of Overton, it defines policy documents even more broadly, simply as "documents written for or by policymakers"[3]. As mentioned above Overton claim to web crawl more than 30,000 national and international sources including governments, intergovernmental organizations (IGOs), think tanks, and charities (Szomszor & Adie, 2022).

Indexing policies

Altmetric.com does not only collect policy document data, but also other altmetrics data like mentions to scientific papers from Twitter or news. In contrast, Overton only focuses on policy documents data, currently indexing more than 6M documents from more than 1500 different policy sources, making it many times larger than similar systems.[4] A much more important difference is that Overton indexes all policy documents and their references, including references to scholarly papers but also references to other policy documents; while Altmetric.com only indexes policy documents explicitly citing scholarly papers. A study from Maleki and Holmberg (2022) compared the policy coverage of 18,996 Scopus publications authored by researchers affiliated with 18 Finnish universities and institutes in eight chosen Social Science fields in Overton and Altmetric.com data sources, checked and compared for coverage and overlap. Their results suggested that on average about 39% of publications were included in Overton and 9% in Altmetric.com. Maleki and Holmberg (2022) results also showed that despite the larger coverage of policy citations in Overton at the institution and field level, Altmetric.com had a tiny but unique proportion of both documents cited (5%) and policy citations that are not covered by Overton (0.6%). There was on average about a 5% overall overlap in the coverage of documents between Overton and Altmetric.com.

There are clearly fundamental conceptual and technical uncertainties on what policy sources Altmetric.com and Overton are currently indexing. It is important to highlight that, different from scientific publications, where "sources" are usually the journals or venues where the scientific publications are published, in the case of policy documents, their "source" is actually the organization that has produced the policy documents. For this reason, in the rest of this paper we will refer to policy sources as *policy organizations*.

2. Objectives

This study addresses these uncertainties by approaching two main exploratory objectives:
(1) to study the overlap and differences in the coverage of policy organizations in Altmetric.com and Overton; and
(2) to compare the distribution and differences of types of policy organizations currently indexed by Altmetric.com and Overton.

With these objectives in mind, we expect to specifically answer the following three research questions:
RQ1: What is the level of consistency of policy organizations covered both by Altmetric.com and Overton?

---

[2] https://help.altmetric.com/support/solutions/articles/6000236695-policy-documents
[3] https://help.overton.io/article/what-sources-does-overton-track/
[4] https://help.overton.io/article/what-is-overtons-coverage-and-how-does-it-compare-to-other-systems/

RQ2: How many of these policy organizations can be found in the open registry ROR[5]?
RQ3: How is the distribution of policy organizations by ROR organization type?

3. Data and methods

3.1. Data collection

We retrieved the policy organization lists from the CWTS in-house versions of the Altmetric.com and Overton databases. The policy organization list from Altmetric.com comprised 471 organizations, while the list from Overton included 1,476 organizations in total. We manually identified those policy organizations that are indexed in both data aggregators.

3.2. Match the data with ROR

To analyze the features of the policy document organizations indexed by both data aggregators, a matching step with a third party was used. The open ROR database has been considered for this matching. By matching policy organizations with the ROR database, we created a common independent framework for studying the types of policy organizations indexed by both aggregators. The ROR database is also available at the CWTS in-house data system. ROR currently contains 103,186 worldwide organizations, including different metadata elements about the organizations like their established year, country, address, and organization type. This element, the organization type[6], is used in this study to analyse the typologies of policy organizations indexes by both Altmetric.com and Overton. We manually matched all the policy organizations reported in both Overton and Altmetric.com with ROR.

4. Preliminary results

4.1. Coverage of policy document source in both database

Figure 1 shows the overlap of coverage of the two databases. It can be seen that only a total of 109 policy organizations are commonly covered by the two aggregators, i.e., 23.1% of policy organizations covered by Altmetric.com are also in Overton, while 7% of policy organizations in Overton are also covered in Altmetric.com. As expected, Overton has a higher overall coverage of sources as compared to Altmetric.com.

Figure 1: Overview of the policy organizations overlap in the two aggregators
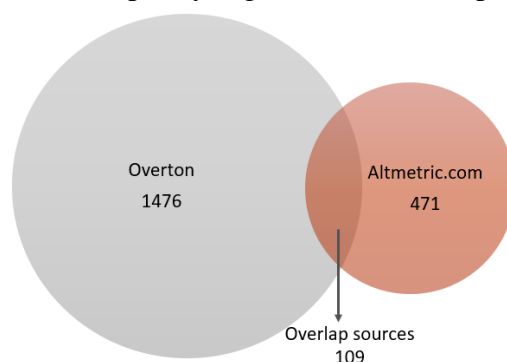
4.2. Coverage of policy organizations in ROR

As shown in Figure 2, we found that 51% (Altmetric.com) and 31% (Overton) of policy organizations are also included in the ROR database. Out of the 109 overlapping policy organizations between Overton and Altmetric.com, 95 (88%) are included in the ROR database.

Figure 2: Overview of the policy document organizations matched with ROR
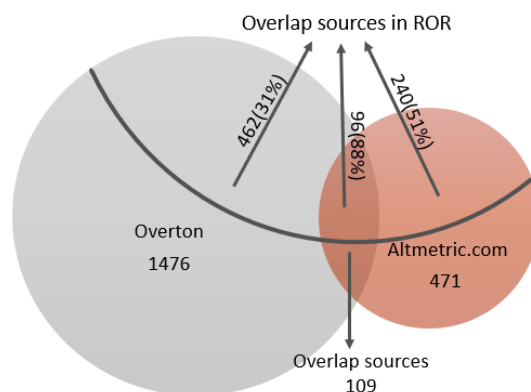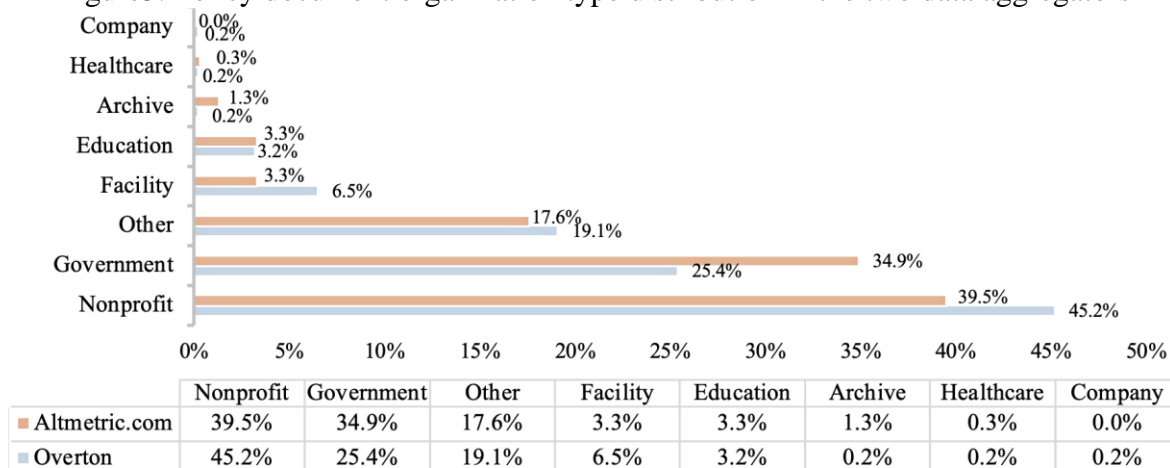

4.3. Distributions of policy organization types

Previous results focus on the coverage and overlap of policy organizations publishing policy documents in general. Another perspective to further investigate the indexing strategies of each data aggregator, and potentially to unveil their underlying definition of policy documents, is to look at the *types* of organizations covered by the two data sources. Figure 3 shows the different policy organization types in Overton and Altmetric.com, as captured in ROR.

In Overton, the majority of policy organizations covered are Non-profit (45.2%), followed by Government (25.4%) and Other (19.1%). The remaining categories, Facility, Education, Archive, Healthcare, and Company, each account for less than 5% of the associated organizations. Altmetric.com follows a very similar pattern, with the largest category of policy organizations being also Non-profit (39.5%), followed by Government (34.9%) and Other (17.6%). The remaining categories also only account for less than 5% of the associated organizations. This suggests that for both Overton and Altmetric.com, Non-profit and Government organizations are the most commonly associated types of policy organizations. Clearly, it will be useful to further investigate why these two categories are the most prevalent and what factors might be contributing to this trend, for example by identifying other subtypes of organizations within these two main types of organizations.

Figure3. Policy document organization type distribution in the two data aggregators

	Nonprofit	Government	Other	Facility	Education	Archive	Healthcare	Company
Altmetric.com	39.5%	34.9%	17.6%	3.3%	3.3%	1.3%	0.3%	0.0%
Overton	45.2%	25.4%	19.1%	6.5%	3.2%	0.2%	0.2%	0.2%

5. Discussion and future work

In 2016, Bornmann et al. (2016) published a paper in which they described policy documents are one of the few altmetrics sources which can be used for target-oriented impact measurement. They said that as a source for the measurement of impact of science on politics, policy documents are one of the most interesting altmetrics sources which should be studied in more detail in future studies. Overton is known for indexing a larger amount of data on policy documents, while Altmetric.com focuses only on policy documents that have cited at least one scientific publication.

To the best of our knowledge, this is one of the first studies to extensively analyze and compare the coverage and distribution of policy document sources (organizations) by both Overton and Altmetric.com. This study corroborates that Overton has a wider coverage of policy organizations than Altmetric.com, since Overton tracks about 1,476 sources as compared to Altmetric.com, which is tracking about 471 sources. However, there are only 109 policy organizations that are tracked by both aggregators. This relatively low overlap between the two data aggregators may suggest that the conceptual discussion about what is a policy document (and what is a policy organization) is a critical one. The fact that two data aggregators aiming at tracking the same object ("policy documents") come up with so different sets of policy organizations, clearly call for more explicit definitions and clarifications of what each of them consider as policy documents, and the motivations for their choices. Researchers interested in the analysis of policy documents should also explicitly consider their definition of policy document, and to implement their data collection based on that definition, rather than on the choices made by the data aggregators.

This study also shows that those policy organizations matched with ROR records are mostly Non-profit and Government organizations, this being consistent among the two data aggregators. This suggests that somehow the two data aggregators have an interest in similar organizations, but somehow have different strategies to identify them. Future research should explore more extensively the different geographies and ages of the organizations covered by both data aggregators.

Finally, this study is also bound by several limitations that need to be acknowledged:
(1) The policy organization matching process between Altmetric.com and Overton was done manually, being a step that can be improved in the future.
(2) The matching process between policy organizations and ROR is highly dependent on the organization name, which should consider the acronym name and translated name of policy organizations, however omissions may still occur. Moreover, ROR is only relevant in a limited manner to study policy organizations, since a substantial number of indexed organizations are not indexed in ROR, which calls for future consideration of dedicated ontologies to better study policy organizations.

References
Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents? *Scientometrics*, 109(3), 1477–1495. Https://doi.org/10.1007/s11192-016-2115-y

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective: Do "Altmetrics" Correlate With Citations? *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. https://doi.org/10. 1002/asi.23309

Fang, Z., Costas, R., Tian, W., Wang, X., & Wouters, P. (2020a). An extensive analysis of the presence of altmetric data for Web of Science publications across subject fields and research topics. *Scientometrics*, *124*(3), 2519-2549.

Fang, Z., Dudek, J., Noyons, E., & Costas, R. (2020b). Science cited in policy documents: Evidence from the Overton database. In *Altmetrics conference*.

Haunschild, Robin; Bornmann, Lutz (2017). "How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data". *Scientometrics*, 110(3), 1209-1216. https://doi.org/10.1007/s11192-016-2237-2

Konkiel, S. (Ed.). (2020). *The State of Altmetrics: A tenth anniversary celebration.* London, UK: Altmetric. https://www.doi.org/10.6084/m9.figshare.13010000

Maleki, A., & Holmberg, K. (2022). Comparing coverage of policy citations to scientific publications in Overton and Altmetric.com: Case study of Finnish research organizations in Social Science. *Informaatiotutkimus*,41(2–3), 92–96. https://doi.org/10.23978/inf.122592

Na, J. C. (2015). User motivations for tweeting research articles: A content analysis approach. *In International Conference on Asian Digital Libraries,* 197-208. Springer, Cham.

Nigel Gilbert, G. (1977). Referencing as persuasion. *Social studies of science*, 7(1), 113-122.

NISO Alternative Assessment Metrics Project. (2014). NISO Altmetrics Standards Project White Paper. Retrieved July 8, 2022, from https://www.niso.org/press-releases/2014/06/niso-issues-altmetrics-white-paper-draft-comment

Pinheiro, H., Vignola-Gagné, E., & Campbell, D. (2021). A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database. *Quantitative Science Studies*, 2(2), 616-642.

Syn, S. Y., & Oh, S. (2015). Why do social network site users share information on Facebook and Twitter? *Journal of Information Science*, 41(5), 553–569. https://doi.org/10.1177/0165551515585717

Szomszor, M., & Adie, E. (2022). Overton: A bibliometric database of policy document citations. *Quantitative Science Studies*, 3(3), 624-650.

Tattersall, A., & Carroll, C. (2018). What can Altmetric. com tell us about policy citations of research? An analysis of Altmetric. com data for research articles from the University of Sheffield. *Frontiers in research metrics and analytics*, 2, 9.