Article

Platforms and Exposure Diversity: Towards a Framework to Assess Policies to Promote Exposure Diversity

Heritiana Ranaivoson * and Nino Domazetovikj

imec-SMIT, Vrije Universiteit Brussel, Belgium

* Corresponding author (hranaivo@vub.be)

Abstract
The fragmentation of consumption and algorithms' increasing impact on how content is recommended and displayed makes it even more important to analyse and promote exposure diversity, i.e., the extent to which audiences are exposed to, discover, and engage with diverse content. Although there is a growing literature addressing how to define media diversity in the context of the challenges posed by platformisation, this article translates the normative dimensions into a framework for operationalising exposure diversity into a tangible policy goal, taking into account datafication and its consequences in terms of increasing data requirements towards platforms. The main objective of this study is to analyse initiatives to assess exposure diversity in the platform era and to discuss how such assessment could be improved, particularly for policy initiatives. This involves addressing several challenges of existing approaches for the assessment of exposure diversity related to defining an appropriate frame of reference, determining the degree of diversity required, dealing with data transparency issues, and promoting user autonomy. To achieve this, we propose a framework for analysing initiatives aimed at assessing and promoting exposure to media diversity. Our framework is composed of four key features: measures (type of initiative), metrics (quantifying exposure diversity), data collection methods, and data requirements. We apply this framework to a set of 13 initiatives and find that policy initiatives can benefit from adopting metrics based on distances and experimenting with data collection methods.

Keywords
audiovisual policy; data requirement; exposure diversity; media pluralism; news policy; online platforms; platform regulation; recommender systems

1. Introduction

With the fast-paced development and adoption of digital technologies in the distribution and consumption of media content, assessing media diversity has become even more important. Media diversity is a significant objective (Helberger, 2011), which is formative for democratic societies. With a long tradition in free speech and democratic theory (Ash, 2016), it has become a central value in media law and policy (Helberger et al., 2020). Having diversity in cultural goods and services offered to citizens, so that they have access to a varied range of available views and ideas, is perceived as a desirable objective for policies addressing the cultural and media sectors. Following Napoli (1997), we distinguish, within media diversity, between source, content, and exposure diversity. Source diversity focuses on content producers and media outlets. Content diversity is concerned with the content features, e.g., the opinions and views expressed, the origin and the language, etc. Exposure diversity corresponds to the extent to which audiences are exposed to, discover, and engage with diverse content (Napoli, 2011).

Due to digitalisation, any content and the ways in which users interact with that content can be transformed into bits of data. Mayer-Schönberger and Cukier (2013) refer to datafication as the process of recording and computing data from different sources and the value that could be obtained from it. Combined, digitalisation and datafication contribute to distinct content types being accessible on the same platform while enabling providers access to unprecedented insights into human behaviour and new forms of creating and extracting value (Kennedy et al., 2015; Mayer-Schönberger & Cukier, 2013). Within this context, the boundaries between media sectors are blurred and there is increased competition between companies that were not competitors before—especially as they now compete for attention (Picard, 2011). Digital transformation and convergence have pushed media companies to extend their set of activities (e.g., newspapers producing podcasts or broadcasters developing on-demand services) resulting in an increase of available content and shifting the competitive landscape from sectoral competition to a situation where many companies and conglomerates have extended their activities beyond the boundaries of their native industries. Due to the proliferation of content across multiple delivery platforms, audiences can conveniently switch between reading news articles on social media and consuming multimedia content on streaming services from a single access point. Digital distribution of content has little effect on costs thus promoting ubiquitous availability across various distribution platforms. Therefore, content convergence can lead to distribution divergence (Dal Zotto & Lugmayr, 2016) as well as consumption fragmentation.

In a context where algorithms modulate how content is recommended and displayed, consumption fragmentation has rendered it more crucial to update the analysis and assessment of exposure diversity (Lambrecht et al., 2022). Considering the changing media landscape, platforms and digital media providers are operating under time and attention constraints, which has highlighted the role of intermediation in creating and capturing value (Kostovska, 2022). These circumstances have prompted the inception of novel approaches to curating and sorting content. There is a complete change in how exposure diversity is considered with the algorithmic influence on recommendation and display (Vrijenhoek et al., 2021) and the possibility to have access to in-depth data on consumption. The core challenge in examining exposure diversity in the platform era is related to algorithms' impact on diversity (Haim et al., 2018; Jürgens & Stark, 2022; Sørensen & Schmidt, 2016). Algorithms can be flawed by inherent biases and can produce unfavourable outcomes such as inequality of representation and coverage (Ranaivoson, 2019), which are associated with challenges identified in the literature as the "filter bubble" (Pariser, 2011) and the "echo chamber" (Colleoni et al., 2014). Colleoni et al. (2014) claim that users are in echo chambers when their prior political views are reinforced

due to selective online exposure to political content. In a similar metaphor, in filter bubbles, users are never aware of what others think outside their bubble regarding political, moral, or scientific issues (Dündar & Ranaivoson, 2022). However, the filter bubble rationale has attracted some criticism. Bruns (2019, p. 8) calls it "the dumbest metaphor on the internet" since, "to the extent that it occurs," homophily results in the first place from users' agency. In addition, empirical evidence that warrants any strong worries about filter bubbles is lacking (Zuiderveen Borgesius et al., 2016). Haim et al. (2018) note that empirical evidence on the existence of the filter bubble and its effects, especially in the context of news, is limited. Of the existing handful of studies, none has been able to prove genuine negative effects of filter bubbles (Ranaivoson, 2019). Nonetheless, the sheer possibility of filter bubbles needs to consider exposure diversity in the context of algorithms.

It is suggested by van Dijck (2014) that it is not accurate to view data as solely a reflection of neutral human behaviour harnessed by platforms and that the role played by platform interventions and interpretations in creating and refining these data resources should be acknowledged. Therefore, established rationales for promoting diversity are being challenged not only by the novel ways of distributing and consuming content but also by the ways in which platforms utilise data to modulate intermediation. As normative notions, media diversity and pluralism are operationalised through measures that aim to increase source and content heterogeneity, with the goal of promoting variety (of information, views, content, and ownership; Ranaivoson, 2007). In contrast, exposure diversity looks at the audience dimension of media diversity (Helberger, 2012). It is therefore not enough to focus on source and content diversity, and it is necessary to consider the specificities of consumption diversity and the role played by all ways of modifying exposure.

Scholarly literature focused on approaches and methodologies for assessing exposure diversity is limited in terms of the volume of studies that incorporate consumption and algorithmic suggestions' diversity. Whereas the issue of defining media diversity in the context of platformisation has been discussed (Helberger, 2018; Helberger et al., 2018; Hendrickx et al., 2020; Joris et al., 2020; Ranaivoson, 2019), the question of how normative dimensions can be translated into tangible measures requires further consideration. Therefore, to advance our understanding of exposure diversity, it is important to adopt a more systematic approach to examine its conceptual features. To accomplish this, we develop a framework for assessing diversity of exposure in the platform era and apply it to a comparison of initiatives aimed at assessing and promoting exposure diversity. Our primary research question is: What characterises initiatives to assess exposure diversity in the platform era and how could such assessment be improved, in particular for policy initiatives?

To answer this question, in Section 2, we discuss the challenges of promoting exposure diversity, which are notably related to the fact that media policy has been mainly aimed at promoting source and content diversity. In Section 3, we present a framework to evaluate initiatives to assess and promote exposure diversity in a datafied society. The following main features of the framework are discussed: measures, metrics, methods, and data requirements. Section 4 presents an overview and compares such initiatives. The main results of the comparison are discussed following our framework. The conclusion focuses on policy implications and recommendations that can be derived from the analysis.

2. The Challenges of Promoting Exposure Diversity

In the EU, media policy has largely sought to promote pluralism by targeting the availability of diverse content. Media diversity has been primarily analysed from the point of view of production and distribution and their impact on consumption. Therefore, current media diversity policies remain rooted in scarcity logic (Helberger, 2019) even though source diversity alone does not secure diversity of the overall output (Helberger, 2012). With digitalisation and datafication, it is important to consider beyond what is available, what actually gets recommended and displayed to users (Ranaivoson, 2020), and the role data play in these processes.

Existing literature highlights several challenges that must be considered to provide a better understanding of the concept. Firstly, as outlined by Helberger et al. (2018), diversity of exposure is a means to an end rather than an aim in itself. Since providing more diversity might come at a cost (Stirling, 2007), it is key to understand how diversity of exposure can contribute to the overall goals. Helberger et al. (2018) highlight three perspectives for considering exposure diversity as a societal goal which can also contribute towards deriving normative positions. These are the individual autonomy perspective, the deliberative perspective, and the adversarial perspective (Helberger et al., 2018). Within this framework, exposure diversity can be utilised to respectively "extend individual choice" and provide individuals with "more opportunities to realize their interests," "promote rational public debate and the formation of a reasoned public opinion," or provide "a corrective to the tendency of public debates to be dominated by existing elites and powerful interests" (Helberger et al., 2018, p. 195).

From a more applied but still conceptual perspective, exposure applies to diverse types and formats of content in different contexts. It seems difficult to design a single approach for assessing and promoting exposure diversity, which would apply to cultural or linguistic diversity or representations of minorities in media outlets. A related critical question from a policy perspective is to decide what an adequate degree of diversity of exposure would be by determining specific benchmarks (Helberger, 2011). There is no consensus on how diversity should be quantified at the recommender system level (Kunaver & Požrl, 2017).

Moreover, difficulties arising from a lack of data transparency and data collection challenges add another layer of complexity to the aspiration of defining indicators to assess exposure diversity. Data are used by media companies to influence consumption behaviours, e.g., in the design and functioning of recommender systems. While such data and data about actual consumption are needed to assess exposure diversity, there is no unified, transparent way for external actors to access it. As proprietary audience data is a valuable resource for developing and maintaining competitive advantage, platforms are incentivised to refrain from data sharing, which can hinder attempts to evaluate exposure diversity (Hagiu & Wright, 2020; van Dijck et al., 2018). Ultimately, data is an asset and firms have an incentive to keep as much as possible of it siloed. For instance, even though Netflix has access to abundant and granular data about the viewing habits and preferences of their users, they only selectively release data on some aspects of their service (Wayne, 2022).

The growing importance of algorithms coupled with the difficulties of predicting their outcomes calls forth for algorithm auditing as an early bias and problem detection tool when new versions are released. However, such an approach can be costly for regulators and those audited and challenging concerning the quantity and quality of data that can be collected as well as the legal soundness of the approach in general (PEReN & Regalia, 2021).

Finally, when discussing the question of data and exposure diversity, a challenge to consider is related to users' agency. Ensuring that users' personal autonomy and privacy and how they envision them are respected (Helberger et al., 2018; Stasi, 2019) can conflict with what experts or policymakers could identify as the "socially-desired distribution of audience attention" (Napoli, 2011, p. 256). Transparency and privacy obligations are at the core of the discussions on designing monitoring mechanisms, as can be seen in the recently adopted EU Regulation Digital Services Act.

3. A Framework to Address Initiatives for Exposure Diversity

Since exposure diversity corresponds to the extent to which audiences are exposed to, discover, and engage with diverse content (Napoli, 2011), it also relates to the users' media diets. Few studies in academic literature incorporate the perspectives of consumption diversity and diversity of algorithmic suggestions when proposing ways for assessing exposure diversity. Furthermore, conceptual advances would be useful for policy purposes.

Therefore, we propose a framework to analyse and compare current or recent initiatives to assess and eventually promote exposure diversity. It distinguishes between measures, metrics, methods, and data

requirements as features of all initiatives included in our overview. Measures are initiatives to achieve more diversity, they correspond to the overall objective, *why* the assessment and promotion of exposure diversity take place. This relates to the discussion on exposure diversity as a societal goal (Helberger et al., 2018). Metrics are less normative and somehow more concrete as they define *what* is being assessed. We expect that the approach changes depending on the type of content and the context. Considering metrics is also rendered necessary by the lack of consensus regarding how to assess exposure diversity—e.g., at the recommender system level (Kunaver & Požrl, 2017). Methods correspond to *how* such assessment is made, notably regarding data collection. Finally, data requirements give one of the main *limits* of the initiatives in terms of Transparency, Parsimony, and Replicability. Actually, while considerable amounts of data are generated, as discussed before, there are large data collection and use challenges.

Our framework first appeared in the context of a study performed for the European Commission on media plurality and diversity online (Lambrecht et al., 2022). It is notably based on a previous study for UNESCO on suitable properties of indexes of media and cultural diversity (Ranaivoson, 2007), which was dealing with *metrics*. The importance of *measures* is highlighted by Helberger et al. (2018), in particular for diversity-by-design measures. The term "framework" itself derives from the Australian Communications and Media Authority (ACMA) study on news measurement framework, which was focused on developing new approaches to assess media diversity in a context of declining local news and a need for better understanding and quantifying public interest journalism. Finally, an original contribution of this framework, compared to the version in Lambrecht et al. (2022), is the emphasis put on data requirements. This is to allow us to take into account how datafication is impacting exposure diversity, i.e., how audiences are exposed to, discover, and engage with diverse content.

3.1. Measures

Measures are initiatives to achieve more diversity. We distinguish between Industry, Policy, and Research initiatives. Industry initiatives are led by one industry partner or a consortium of industry partners to develop solutions to improve exposure diversity. Policy initiatives include law-making and policies aimed at monitoring or improving exposure diversity. Research initiatives are led by universities or academic institutions to assess exposure diversity. Among the latter, we have tried to exclude Research projects of a smaller scale. There are indeed a great number of studies that aim at increasing exposure diversity—see, e.g., Latha and Nadarajan's (2019) mapping of approaches to improve recommendation diversity). They can be focused on software solutions. For example, Kamishima et al.'s (2012) recommender system aims at providing neutral recommenda-

tions to users in relation to a specific viewpoint. Latha and Nadarajan's (2019) approach incorporates diversity into its recommendations and is applied to movies and news. Helberger (2018) recommends the use of tools for alternative recommendation settings and technologies that make users aware of their filter bubble.

3.2. Metrics

Metrics refer to what is being assessed. This includes of course assessing the diversity of exposure, but also identifying the barriers to diversity. Recent literature overviews point to the too high number of conceptualisations of news and media diversity (Hendrickx et al., 2020; Joris et al., 2020). Besides, media diversity is assessed by a broad spectrum of scholars, from social to computer sciences. Despite this, the initiatives to promote and assess exposure diversity use indicators that derive from and sometimes combine the four following basic metrics: Count, Percentages, Dual Indexes, and Distances (Lambrecht et al., 2022). These basic metrics can be aggregated or combined to assess the diversity of exposure.

Count represents the number of elements that are part of a set. The elements can be various units, which can be part of a subset or grouped in a category. Count can be used to quantify to what extent units of various subsets or categories are represented in a set. Percentages correspond to all indicators consisting in measuring the relative share of a category within a system, e.g., the shares of films in Netflix's catalogue per country of origin. Dual Indexes (Stirling, 2007) notably include both the entropy index and the Herfindahl-Hirschman index. As Stirling's (2007) denomination alludes, both indices combine in their quantification what Count and Percentages assess respectively.

Distances correspond to the level of differences, or dissimilarity, between every pair of items consumed, for example. Distances are commonly used by computer scientists to assess diversity (Kunaver & Požrl, 2017). Using cosine similarity, they assess how similar two items are. Mathematically, cosine similarity corresponds to the normalised (between 0 for *maximum similarity* and 1 for *maximum dissimilarity*) cosine of the angle between two vectors that are projected in a multidimensional space (Prabhakaran, 2018). Distances can be calculated for any content, assuming that the data is labelled (e.g., using metadata) or can be categorised (Lambrecht et al., 2022). For example, a common computational approach to assess distances is the bag-of-words model, which compares texts according to the words they use. Distances are often used to assess recommender systems' diversity—especially in comparison to Count or Percentages. In practice, the diversity of recommendations will be increased by suggesting items that are further from a user's preferences, although always within certain limits (Lambrecht et al., 2022). For example, someone who has watched a video about football

could be proposed other videos about other sports, because football is a subcategory of sport or because the recommender system has found that those who have watched videos about football often also watch videos about other sports.

3.3. Methods

Methods are ways to collect data that can be used to assess the diversity of exposure. Methods include Surveys, Online Data Scraping, and Experiments. Surveys consist in gathering information from a sample of individuals, usually through a questionnaire but this also includes interviews. An example is the Office of Communications's (Ofcom) collection of data via consumer research for their Measurement Framework for Media Plurality (MFMP). Collected information is therefore declarative. In contrast, Online Data Scraping consists in extracting data from human-readable output or content available online, which provides insights into actual (consumption) behaviour at aggregate and, in some cases, individual levels. For example, for the reports on "Australian content," the researchers entered into a database all titles identified on Stan, Netflix, and Amazon Prime Video when inspecting content included in relevant categories on each service (such as Australian movies and Australian TV) and using search terms ("Australia," "Australian," "Austral*") to identify relevant content not included in these categories (Lobato & Scarlata, 2019). Experiments also aim at analysing the actual behaviour of individuals but by putting them in a setting where their behaviours can be directly observed and hence more easily explained. Experiments can also consist of simulations to test features such as algorithmic systems. For example, in the discoverability index Research initiative, Tétu and Dubois-Paradis (2020) built two profiles on Netflix to observe which films get recommended to these profiles, based on their different viewing behaviours.

3.4. Data Requirements

Finally, the initiatives can be compared based on data requirements. The emphasis is put upon which data would be required for the methodologies to be applied. The fourth feature of our framework considers the profound changes in media consumption resulting from datafication. Huge amounts of real-time data are produced but this is also challenging for initiatives aiming at promoting and/or assessing exposure diversity: Given the available data, a suitable initiative should not be too demanding in terms of data necessary to build metrics. Within this feature, we distinguish between Transparency, Parsimony, and Replicability. *Transparency* means how easy the measures, metrics, and methods are to understand. The initiative is Transparent when the assumptions are explicit, including regarding the data used to compute the metrics and more generally apply

the initiatives. *Parsimony* corresponds to the costs underlying the use of data by the initiatives. This includes the financial cost to access the data and, beyond, how simple it is to assess exposure diversity (e.g., required computing power and conversely the speed at which calculations can be done). *Replicability* means to what extent the indicators can repeatedly be assessed over time in the same context. This notably concerns the type of data the frameworks rely on. Questions here include the type of data, their periodicity, and how they are collected. Besides, the access to data is considered: To what extent are data available publicly? Who owns such data? How is access to these data ensured? And to whom?

4. A Review of Initiatives for Exposure Diversity

4.1. Methodology

We employ a multi-method approach to investigate a nascent aspect of media and communications policy including document analysis (Karppinen & Moe, 2012) and expert interviews (Van Audenhove & Donders, 2019). Expert interviews are suitable for gathering non-codified knowledge and insights about emerging issues in media and communications policy research, whereas the selection of interviewees is identified as a critical aspect of the research design (Hammersley & Atkinson, 1995). We conducted 15 semi-structured expert interviews with professionals, academics, and technical experts from various geographic locations such as Belgium, the Netherlands, France, the UK, Australia, Canada, and Switzerland (see Table 1). A snowball effect was created by encouraging the interviewees to suggest other potentially relevant interviewees. The information collected through the interviews allowed us to complement and contextualise the document analysis of 13 initiatives aimed at assessing exposure diversity (see Table 2). The selection of initiatives was based on a snowball process as well. The document analysis involved a systematic comparison of the initiatives, with the coding being informed by both publicly available information and undisclosed information gathered during the interviews. The framework used for the analysis enabled the initiatives to be compared in terms of why, what, how, and to what extent they allow for the assessment of exposure diversity.

Regarding measures, initiatives can be either Policy, Industry, or Research. Metrics can include and sometimes cumulate Count, Percentages, Distances, and Dual (for Dual Indexes). Similarly, there can be one or more methods among Surveys, Scraping, and Experiments. Data requirements for each of these initiatives are more difficult to assess in a comparative manner. This is why we propose a more granular approach and distinguish between Transparency, Parsimony, and Replicability. The assessment though remains a bit subjective and qualitative, especially in comparison to our other features. Transparency and Replicability refer to easiness, respectively to understand and to repeat, which are difficult

Table 1. Interviews conducted in the frame of the project.

Name	Position	Organisation	Date of interview
Glen Joris	PhD student	Ghent University	30/11/2020
Joris Mattheijssens	Data scientist	Vlaamse Radio- en Televisieomroeporganisatie	30/11/2020
Sanne Vrijenhoek	Researcher	University of Amsterdam	02/12/2020
Johan Loeckx	Researcher	Vrije Universiteit Brussel	02/12/2020
Nicolas Rolin Lucas Verney Benoît Rottembourg	Researchers	Pôle d'Expertise de la Régulation Numérique Inria	03/12/2020 16/02/2021
Nava Tintarev	Professor	TU Delft	04/12/2020
Olaf Steenfadt	Global project director	Media Ownership Monitor	10/12/2020
Michèle Rioux	Director	Université du Québec à Montreal	10/12/2020
Eleonora Mazzoli	PhD student	London School of Economics and Political Science	08/01/2021
Ramon Lobato	Professor	Royal Melbourne Institute of Technology	21/01/2021
Nanao Kachi	Director of social and consumer policy	Canadian Radio-Television and Telecommunications Commission	22/01/2021
Véronique Guèvremont	Professor	Laval University	22/01/2021
Sébastien Noir	Head of software engineering, technology, and innovation	European Broadcasting Union	25/01/2021
Catherine Johnson	Professor	University of Huddersfield	27/01/2021

to objectively assess. Parsimony refers to costs but all data was not available to extensively compare initiatives on that aspect. For all these reasons, that feature of the framework should be taken cautiously. Similarly, the information through our document analysis and our interviews could not always allow us to properly assess this feature, hence the note that data is in that case not available (n/a).

4.2. The Rise of Policy Initiatives Towards Exposure Diversity

Considering measures, the comparison first shows that Policy initiatives have appeared that take exposure diversity into account. Napoli (2011) deemed that exposure diversity, despite being increasingly regarded as a critical component when assessing diversity within online environments, is very seldom the object of policy interventions. This seems to be changing now. While media diversity remains usually assessed by regulatory authorities by considering source diversity (ACMA, 2020), our sample includes four Policy initiatives: two established ones (Ofcom's MFMP and the EAO's yearly publication on the visibility of audiovisual works on transactional video-on-demand services) and two more in the making. For the latter, two reports published respectively in Australia (ACMA, 2020) and in Canada (Parliament of Canada, 2021) could lead to policies to promote exposure diversity among others. Bill C-11 is the second attempt to amend Canada's Broadcasting Act after a bid

to modernise the Act with Bill C-10 ended unsuccessfully with the dissolution of the 43rd Canadian Parliament in 2021.

Besides one Industry initiative, most other initiatives are Research initiatives. As algorithms (notably recommender systems) and their impact on diversity play a significant role in the analysis of exposure diversity, most research revolves around computer science. However, some projects like PersoNews or DIAMOND rely on an interdisciplinary approach, involving social sciences, e.g., to analyse the impact of algorithms on users, to understand how they fit in media business strategies, or to devise policy recommendations.

4.3. A Focus on One or Two Metrics

All metrics appear in our sample of initiatives: Percentages in eight of them, Distances in seven of them, Count in six of them, and Dual in three of them. Only two Research initiatives rely on all these metrics, PersoNews and ENSURE, which reflects the broad scope of research they involve. Both include several more focused projects, which can have different emphases in terms of exposure diversity.

Another interesting specificity concerns the use of Distances, which has so far never been used in Policy initiatives but exclusively in Research initiatives (the European Broadcasting Union PEACH Industry initiative being the exception). Moreover, these Research initiatives always involve computer scientists, which reflects

Table 2. A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
EAO Visibility	Audiovisual	Policy	Percentages and Count	Scraping	n/a	Yearly analysis by the EAO (European Audiovisual Observatory) of the visibility (promotion) of audiovisual works on transactional video-on-demand services	EAO	Europe
Australian content	Audiovisual	Research	Count and Percentages	Scraping	Transparent and Replicable	Yearly reports examining the availability and discoverability of Australian screen content on subscription video-on-demand services	Royal Melbourne Institute of Technology	Australia
Broadcasting Act	Broadcasting	Policy	n/a	n/a	n/a	A 2020 report reviewed the regulatory and legislative frameworks for broadcasting and telecommunications in Canada; the legislation (Bill C-11, currently under final consideration in the Canadian Senate) could lead to some recommendations being applied	Canadian Radio-Television and Telecommunications Commission	Canada
PEACH	Broadcasting	Industry	Distances	Scraping	Transparent and Parsimonious	Personalisation for EACH (PEACH) is a "personalisation and recommendation ecosystem developed by broadcasters for broadcasters"; the recommended content should broaden a user's horizon	European Broadcasting Union	Switzerland
Discoverability index	Music Audiovisual Books	Research	Percentages	Experiments	Transparent and Replicable	Index considering the presence, visibility, and recommendation of content (music, audiovisual, book), developed by the LATICCE lab	Université du Québec à Montréal	Canada
Recoloco	n/a	Research	Distances	Scraping	n/a	REcommanding personalized COntent for LOcal Communities (Recoloco) has developed software and approaches to personalise Postbuzz's content and user experience, a digital replica of your physical mailbox; this has included automatically identifying new content tags, recommending content tags, and profiling users	Vrije Universiteit Brussel	Belgium

Table 2. (Cont.) A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
MFMP	News	Policy	Count, Percentages, and Dual Indexes (Herfindahl-Hirschman index)	Scraping and Surveys	Transparent	The MFMP was developed by the UK Ofcom to measure media diversity via availability, consumption, impact, and contextual factors	Ofcom	UK
PersoNews	News	Research	All	Scraping and Experiments	n/a	The European Research Council PersoNews (profiling and targeting news readers and implications for the democratic role of the digital media, user rights, and public information policy) project investigated the impact the trend for personalisation has on the role of digital media in society and how that can be assessed; it has been followed by several projects such as the Research Priority Area Human(e) AI	University of Amsterdam	Netherlands
NewsDNA	News	Research	Percentages and Distances	Experiments	Transparent and Replicable	Diversity in the News Through Algorithmization (NewsDNA) was an interdisciplinary four-year research project (2018–2022) to develop and test an algorithm that uses news diversity as a key driver for personalised news recommendations	Ghent University	Belgium
DIAMOND	News	Research	Distances	Surveys	Transparent and Replicable	Diversity and Information Media: New Tools for a Multifaceted Public Debate (DIAMOND) was a Flemish (Belgian) interdisciplinary four-year research and valorisation project (2017–2021) on news diversity	KU Leuven	Belgium

Table 2. (Cont.) A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
CPN	News	Research	Distances	Surveys and Experiments	Transparent and Parsimonious	The EU H2020 Content Personalisation Network (CPN)project has offered news organisations transparent and easily integrated software to personalise their content ("bring your audience the right stories at the right time'')	Vlaamse Radio- en Televisieomroe-porganisatie	Belgium
ACMA report	News	Policy	Count and Percentages	Survey and Scraping	Transparent and Replicable	The ACMA 2020 report (*News in Australia: Diversity and Localism. News Measurement Framework*) sets out an alternative framework to measure the current levels of news diversity and the availability of local news throughout Australia	ACMA	Australia
ENSURE	Various	Research	All	Surveys and Experiments	Transparent and Replicable	The ExplaiNing SeqUences in REcommendations (ENSURE) project looked at ways of improving the transparency and decision support for recommender systems; it is linked to other research projects on diversity (recommender systems, viewpoint diversity, etc.)	TU Delft	Netherlands

the fact that computer scientists tend to assess diversity by using Distances (Kunaver & Požrl, 2017).

Conversely, initiatives that make use of Count and/or Percentages seldom make use of Distances. The problem is that these metrics consider different, complementary aspects of diversity, as the Stirling (2007) model shows. For example, increasing the diversity of recommendations on a music streaming platform can be done by proposing songs by more different artists, instead of always by the same artist (which can be assessed by combining Count and Percentage) as well as by proposing songs in less familiar genres (which can be assessed in terms of Distances).

4.4. Methods: Main Results

The three types of methods (Surveys, Scraping, and Experiments) appear well-represented in our sample. The most striking result is the fact that Policy initiatives never rely on Experiments. One explanation could be that such Experiments better fit exploratory analysis whose results still need to be generalised after, which makes them probably less fit in the context of a Policy initiative.

The analysis of the initiatives also shows that most Research initiatives on exposure diversity involve computer science. This seems logical in a context where algorithms (notably recommender systems) and their impact on diversity play a significant role in exposure diversity. Some Research initiatives (and in our view, the most interesting ones) rely on an interdisciplinary approach, involving social sciences, e.g., to analyse the impact of algorithms on users, to understand how they fit in media business strategies, or to devise policy recommendations.

4.5. Transparent but not Parsimonious Initiatives

Finally, regarding data requirements, most initiatives are Transparent (nine vs. three). The assessment was made based on the information provided about data used to assess and eventually improve exposure diversity, as well as on the metrics actually used. Neither Recoloco nor EAO Visibility provided information on the data used (Recoloco notably provided no information on the data used to profile users). As for PersoNews, the lack of Transparency rather stemmed from the lack of information on the metrics themselves.

Only two initiatives are Parsimonious: PEACH and CPN. PEACH provides tools aimed at editors and journalists. It is developing algorithms to recommend content which will broaden a user's horizon (i.e., to educate them). Data used are the ones collected already by media providers, it is rather the functioning of the algorithms that is tweaked in another direction. In the CPN Research initiative, exposure diversity was assessed based on partner public service media's algorithmic systems and a rather classical approach, hence Parsimony,

although user research is more costly in the process. Interestingly, a common point of PEACH and CPN is that they are led by media service providers.

Finally, six initiatives are Replicable. Replicability depends on data type and ownership. A difference is to be made notably between whether data are available publicly or not. Hence, EAO Visibility is relying on a partnership with Ampere Analysis, the EAO is therefore not owning data. In comparison, for the discoverability index, the LATICCE team has set up profiles on streaming services to compare received recommendations, which could be replicated by anyone with internet access and a subscription to the analysed services.

5. Discussion

The increasing importance of platforms in modern society has contributed to a growing apprehension from governments and advocacy groups regarding issues such as privacy and surveillance concerns, potential threats to freedom of expression, as well as the possibility of technological and infrastructure domination (Gillespie, 2018). Data is increasingly important but is often controlled by a few powerful actors. The resulting imbalance of power between those who provide the data and those who control it can impact how institutions and public discussions are governed (Kennedy et al., 2015; van Dijck et al., 2018). Diversity, as a crucial objective of media policy, needs to be readdressed in a context where online platforms' domination over access to content has altered the way citizens are exposed to media. The aim of this article is to provide a framework to analyse initiatives meant to assess and promote exposure diversity in the platform context. We develop and apply such analytical framework to a sample of initiatives and highlight drawbacks as to the preparedness of current policy in considering datafication when assessing exposure diversity across media landscapes. We highlight the importance of considering exposure diversity as a policy goal and propose approaches for defining and assessing it.

Our main contribution is the framework to assess and compare initiatives towards exposure diversity. This framework consists of four features: measures, metrics, methods and data requirements. Measures are initiatives to achieve more diversity and they correspond to the overall objective, why the assessment and promotion of exposure diversity take place. Metrics are less normative and somehow more concrete as they define what is being assessed. Methods correspond to how such assessment is made, notably regarding data collection. Finally, data requirements consider the changes that datafication has brought to the whole media consumption process, assessing the initiatives in terms of Transparency, Parsimony, and Replicability. The framework is applied to a set of 13 initiatives aimed at assessing and eventually promoting exposure diversity. Applying the framework allows us to compare in a systematic way these initiatives according to why, what,

how, and to what extent these initiatives allow us to assess exposure diversity. Besides, the features of the frameworks can be combined in the analysis.

All this allows us to respond to the research question: What characterises initiatives to assess exposure diversity in the platform era? How could such assessment be improved, in particular for Policy initiatives?

Firstly, as exposure diversity is recognised in the policy sphere as an important objective that requires policies to promote and foster it, there is a growing number of Policy initiatives aimed at assessing and promoting exposure diversity. Secondly, based on our analysis, it may be argued that future designs of policy measures need to include Distances as metrics, as direct applications of research (in computer science). The focus on Percentages and Count leads to a reductive view of only one side of the problem. Distances have been applied only rarely as metrics to assess diversity in the cultural and media sectors (for example, see Farchy & Ranaivoson, 2011). One reason may be that measuring Distances has proved difficult for a long time for social scientists, either conceptually or technically. However, as computer scientists are doing it on a regular basis, there should be fewer obstacles now to such an endeavour. Furthermore, we recommend more generally for future Policy initiatives to develop a set of indicators instead of very synthetic indexes. This would enable adopting a multifaceted approach in grasping challenges related to diversity and a more comprehensive operationalisation of diversity.

Thirdly, Policy initiatives never rely on Experiments. One explanation could be that such Experiments are better suited for exploratory analysis and their results still need to be generalised after, which likely makes them less fitting in a policy context. More general considerations in terms of methods are about the importance of an interdisciplinary perspective for Policy initiatives and more generally for all initiatives. Social sciences and computer sciences remain too separated while some of the most interesting frameworks rely on interdisciplinary approaches. In a similar way, it is also crucial to involve experts and practitioners in the design and piloting of measurement frameworks, such as media professionals, algorithm developers, representatives from non-profit organisations and civil society, etc. Fourthly, for Policy initiatives as for other initiatives, it is important to assess constraints posed by data requirements. Transparency, Parsimony, and Replicability are important for citizens to understand these Policy initiatives and for such initiatives to be efficient and applied in the long term.

6. Conclusion

The proposed framework to analyse initiatives aimed at assessing and promoting exposure diversity in the platform context could also be used to assess future initiatives. Several important measurement frameworks are under discussion and should be followed up. In Canada,

the proposed Bill C-11 is under discussion that would regulate the outcome of recommender systems of platforms and digital media providers serving the Canadian market with the aim to improve the prominence and discoverability of local content, which could likely lead to adopting new or adapting existing measurement frameworks. In Australia, the ACMA's report would still need to be validated and implemented before becoming an actual measurement framework.

A major difficulty lies in how to best involve—or address—online platforms. Online platforms impact all media sectors through algorithmic gatekeeping (Helberger, 2019; Napoli, 2015), which they deploy in order to automatically filter, rank, and recommend content (Haim et al., 2018). Nonetheless, understanding their impact necessitates an ability to evaluate the scope of diversity they make available and recommend. Currently under debate is whether platforms should be obliged to share data with national regulatory authorities, given the risks related to technological and infrastructure domination and the fundamental imbalances between those that provide data and those who control it (Gillespie, 2018; Kennedy et al., 2015; van Dijck et al., 2018). Alternatively, researchers could be allowed to employ data-gathering methods that do not require the platforms' authorisation—e.g., by recruitment of internet users to install an apparatus that automatically records and reports their internet usage (see Kitchens et al., 2020) or by using bots to scrape and collect data. This could benefit from online platforms being incentivised to share data allowing to measure media diversity. In any case, doubts would remain regarding the collected data's reliability and the metrics to be devised to audit platforms.

A follow-up and more extensive overview of initiatives used to not only measure exposure diversity but, beyond, promote prominence, discoverability, and serendipity, will be key to ensuring initiatives to assess and eventually promote exposure diversity remain relevant.

Acknowledgments

Conflict of Interests

The authors declare no conflict of interests.

References

Ash, T. G. (2016). *Free speech*. Yale University Press.
Australian Communications and Media Authority. (2020). *News in Australia: Diversity and localism*. https://www.acma.gov.au/publications/2020-12/report/news-australia-diversity-and-localism
Bruns, A. (2019). Filter bubble. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1426

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317–332.

Dal Zotto, C., & Lugmayr, A. (2016). Media convergence as evolutionary process. In A. Lugmayr & C. Dal Zotto (Eds.), *Media convergence handbook: Firms and user perspectives* (Vol. 2, pp. 3–16). Springer.

Dündar, P., & Ranaivoson, H. R. (2022). Science by YouTube: An analysis of YouTube's recommendations on the climate change issue. *Observatorio (OBS*)*, *16*(3). https://doi.org/10.15847/obsobs16320222061

Farchy, J., & Ranaivoson, H. (2011). Do public television channels provide more diversity than private ones? *ENCATC Journal of Cultural Management and Policy*, *1*, 50–63.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Hagiu, A., & Wright, J. (2020). When data creates competitive advantage. *Harvard Business Review*, *98*(1), 94–101.

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330–343.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. Routledge.

Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, *1*, 441–469.

Helberger, N. (2012). Exposure diversity as a policy goal. *Journal of Media Law*, *4*(1), 65–92.

Helberger, N. (2018). Challenging diversity—Social media platforms and a new conception of media diversity. In M. Moore & D. Tambini (Eds.), *Digital dominance: The power of Google, Amazon, Facebook, and Apple* (pp. 153–175). Oxford University Press.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207.

Helberger, N., Moeller, J., & Vrijenhoek, S. (2020). *Design pour la diversité—Diversité des contenus à l'ère numérique* [Design for diversity—Content diversity in the digital era]. Government of Canada. https://www.canada.ca/fr/patrimoine-canadien/services/diversite-contenus-ere-numerique/design-diversite.html

Hendrickx, J., Ballon, P., & Ranaivoson, H. (2020). Dissecting news diversity: An integrated conceptual framework. *Journalism*, *23*(8), 1751–1769.

Joris, G., De Grove, F., Van Damme, K., & De Marez, L. (2020). News diversity reconsidered: A systematic literature review unraveling the diversity in conceptualizations. *Journalism Studies*, *21*(13), 1893–1912.

Jürgens, P., & Stark, B. (2022). Mapping exposure diversity: The divergent effects of algorithmic curation on news consumption. *Journal of Communication*, *72*(3), 322–344.

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Enhancement of the neutrality in recommendation. In M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, G. Semeraro, & M. Willemsen (Eds.), *Decisions@RecSys'12* (pp. 8–14).

Karppinen, K., & Moe, H. (2012). What we talk about when we talk about document analysis. In M. Puppis & N. Just (Eds.), *Trends in communication policy research: New theories, methods and subjects* (pp. 177–193). Intellect.

Kennedy, H., Poell, T., & van Dijck, J. (2015). Data and agency. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715621569

Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS Quarterly*, *44*(4), 1619–1649. https://doi.org/10.25300/misq/2020/16371

Kostovska, I. (2022). Blockchain ecosystems in the creative industries: Big dreams for micro-payments. In S. Baumann (Ed.), *Handbook on digital business ecosystems* (pp. 405–423). Edward Elgar.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems—A survey. *Knowledge-Based Systems*, *123*, 154–162.

Lambrecht, I., Mazzoli, E. M., Ranaivoson, H., Domazetovikj, N., Tambini, D., & Valcke, P. (2022). An assessment of the effectiveness of measures related to prominence and the findability/discoverability of general interest content and services. In Directorate-General for Communications Networks, Content and Technology (Ed.), *Study on media plurality and diversity online: Final report* (pp. 95–157). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/475bacb6-34a2-11ed-8b77-01aa75ed71a1/language-en/format-PDF/source-266738523

Latha, R., & Nadarajan, R. (2019). Analysing exposure diversity in collaborative recommender systems—Entropy fusion approach. *Physica A: Statistical Mechanics and its Applications*, *533*, Article 122052.

Lobato, R., & Scarlata, A. (2019). Australian content in SVOD catalogs: Availability and discoverability. Digital etnography Research Centre.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Napoli, P. M. (1997). Rethinking program diversity assessment: An audience-centered approach. *Journal of Media Economics*, *10*(4), 59–74.

Napoli, P. M. (2011). Exposure diversity reconsidered. *Journal of Information Policy*, *1*, 246–259.

Napoli, P. M. (2015). Social media and the public inter-

est: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751–760.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.

Parliament of Canada. (2021). *C-11: An act to amend the Broadcasting Act and to make related and consequential amendments to other Acts*. https://www.parl.ca/legisinfo/en/bill/44-1/c-11

PEReN, & Regalia. (2021). *Evaluation methods for content recommendation algorithms*. Direction Générale des Médias et des Industries Culturelles. https://www.peren.gouv.fr/rapports/2021-05-06%20-%20Evaluation%20Methods%20for%20Content%20Recommendation%20Algorithms%20-%20PEReN-Regalia.pdf

Picard, R. G. (2011). *The economics and financing of media companies*. Fordham Univ Press.

Prabhakaran, S. (2018). *Cosine similarity—Understanding the math and how it works?* machinelearningplus. https://www.machinelearningplus.com/nlp/cosine-similarity

Ranaivoson, H. (2007). *Measuring cultural diversity: A review of existing definitions*. UNESCO.

Ranaivoson, H. (2019). Online platforms and cultural diversity in the audiovisual sectors: A combined look at concentration and algorithms. In L. A. Albornoz & M. T. Garcia Leiva (Eds.), *Audiovisual industries and diversity economics and policies in the digital era* (pp. 100–118). Routledge.

Ranaivoson, H. (2020). Cultural diversity. In R. Towse & T. Navarrete Hernández (Eds.), *Handbook of cultural economics* (3rd ed., pp. 183–191). Edward Elgar.

Sørensen, J. K., & Schmidt, J. H. (2016). *An algorithmic diversity diet? Questioning assumptions behind a diversity recommendation system for PSM*. Unpublished manuscript.

Stasi, M. L. (2019). Social media platforms and content exposure: How to restore users' control. *Competition and Regulation in Network Industries*, *20*(1), 86–110.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719.

Tétu, M., & Dubois-Paradis, S. (2020). *Quelle place pour les produits audiovisuels québécois en ligne? Un modèle d'observation empirique de la découvrabilité* [What place for Quebecois audiovisual products online? An empirical observation model of discoverability]. LATTICE; CEIM.

Van Audenhove, L., & Donders, K. (2019). Talking to people III: Expert interviews and elite interviews. In H. Van den Bulck, M. Puppis, K. Donders, & L. Van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 179–197). Palgrave Macmillan.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.

van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., & Helberger, N. (2021). Recommenders with a mission: Assessing diversity in news recommendations. In F. Scholer & P. Thomas (Eds.), *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 173–183). Association for Computing Machinery.

Wayne, M. L. (2022). Netflix audience data, streaming industry discourse, and the emerging realities of "popular" television. *Media, Culture & Society*, *44*(2), 193–209.

Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vresse, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). https://doi.org/10.14763/2016.1.401

About the Authors

Heritiana Ranaivoson is a senior researcher and project leader at Vrije Universiteit Brussel's imec-SMIT (Belgium) He holds an MSc in economics and management from the École Normale Supérieure de Cachan and a PhD in industrial economics from Université Paris 1, Panthéon-Sorbonne. He has led several research projects at international and national levels for the European Commission (H2020, study contracts), Unesco, Google, and others. He has published extensively in the fields of cultural and media diversity, media policy, audiovisual platforms, and recommender systems.

Nino Domazetovikj is a PhD researcher at Vrije Universiteit Brussel's imec-SMIT (Belgium) where he is part of the Media Economics and Policy unit. His research interests are situated at the crossroads of media economics, media management, and media policy studies. His current work is focused on the operations of global video-on-demand providers and how they alter financing and production practices in European audiovisual markets.