Methodological challenges in the transition towards online audience research

David Mathieu, Roskilde University, Denmark
Miguel Vicente-Mariño, Universidad de Valladolid, Spain
Maria José Brites, Lusophone University of Porto and Communication and Society Research Centre, Portugal
Inês Amaral, Universidade Autónoma de Lisboa, Portugal
Niklas A. Chimirri, Roskilde University, Denmark

Juliane Finger, Universität Hamburg, Germany
Bojana Romic, Independent Researcher, Denmark
Minna Saariketo, University of Tampere, Finland
Riitta Tammi, Aalto University, Finland
Marisa Torres da Silva, New University of Lisbon, Portugal
Liliana Pacheco, Lisbon University Institute, Portugal

(With a contribution from Félix Ortega, University of Salamanca, Spain)

Abstract:
This review of the literature published between 2005 and 2014 presents an overview of the methodological environment in which audience research is transiting towards the study of online audiences. Online audience research is a mix of long-established research rationales,
methodical adaptations, new venues and convergent thinking. We discuss four interconnected, and sometimes contradictory, methodological trends that characterize this current environment: 1) the expansion of online ethnography and the continued importance of contextualization, 2) the influence of big data and an emphasis on uses, 3) the reliance on mixed methods and the convergence of different rationales of research, and 4) the ambiguous nature of online data and the ethical considerations for the conduct of research. In spite of a massive research activity, there remain gaps and underprivileged areas that call for a re-prioritization of research. In the conclusion of this paper, we offer recommendations to orient future research.

**Keywords:** Online Audience, New Media, Research Method, Methodology, Literature Review, Big Data, Ethnography, Contextualization, Ethics, Mixed Method, Convergence.

Given the recent transformations in the media landscape, much audience research has turned to the study of new, digital and social media (see Zeller et al., 2015 or Bredl, Hünniger & Linaa Jensen, 2014). Online presence is bringing new challenges and opportunities to both scholarly and commercial research (Vicente-Mariño, 2013), ranging from the growing amount of available data to the depth of information potentially accessible. The purpose of this paper is to report on the methodological challenges and opportunities facing audience research in its transition towards the study of online audiences as a response to a changing media environment.

The dynamics of media production and consumption have changed substantially via processes of convergence, digitalization and the development of the internet, creating new audience practices oriented towards interaction (Schmidt, 2013), participation (Carpentier, 2011), produsage (Bruns & Jacobs, 2006; see also Pavlickova & Kleut, this issue) or globalization (Mackay & Tong, 2010). As a result of these changes, audience engagement has shifted both qualitatively and quantitatively. Terms such as ‘digital natives’ (Prensky, 2001) or the ‘net generation’ (Tapscott, 1998), although contested (Helsper & Eynon, 2009), do indicate that young people are able to navigate fluidly and shift easily between online and offline environments (Wang, 2012, Wang, 2010) and to use social media for information, communication and expression needs (Mihailidis, 2014), to the point where structural categories that traditionally organised media consumption seem to lose their relevance online (Hermes, 2009).

This new media environment has triggered debates whether new methods and methodologies are needed (Gauntlett, 2009; Merrin, 2009). Research has been quick to take advantage of the new affordances of technology, and much has been written on the study of online data and big data. But have these developments really changed the scientific requirements of audience research, for example regarding validity, ethics or contextualization? Indeed, some in the field have advocated continuity, rather than rupture (Carpentier, 2011; Press & Livingstone, 2006; Livingstone, 2004). Whereas the technological
innovation in the media sphere happens quickly, the evaluation and methodological transition requires more time in order to meet the quality requirements of scientific inquiry. The fast technological evolution can be seen not only as opportunity to renew our observational and analytical resources and devices, but also as a valuable chance to further explore and renegotiate our connections with the classical repertoire of audience research procedures.

Therefore, we see a timely need for tracing the transition from traditional (or what could be more appropriately called ‘pre-online’) audience research towards the study of online audience practices. Not only will such a literature review help in establishing the points of continuity and rupture within the tradition of audience research, but it will also allow assessing the direction that research is currently taking in a research environment that struggles to cope with the pace with which the media landscape is changing.

Methodological considerations
This paper reviews the literature on online audience research over a decade spanning from 2005 to 2014, with the aim of portraying the methodological challenges and opportunities facing online audience research as well as identifying gaps and milestones that have occurred during this transitional period. We assume the term online audience research to be a rather uncontroversial one, referring to the study of audiences’ material and symbolic practices in online environments. While such a definition can give rise to difficulties of application in some circumstances (conventional television is now connected to the internet, people use applications or software that can be switched on and off), we believe a central aspect of online audience research is related to the presence or identities that people assume online.

One of the main challenges we faced was to identify relevant literature in audience research (whether it could be labelled ‘online’ or not). With the increased popularity of new and social media, a massive amount of work is being produced representing all walks of research. Besides media and communication studies, which are well represented, we have encountered arguably relevant work in sociology, anthropology, political sciences, computer sciences and technology studies, research of both the qualitative and quantitative types and which relies on a multitude of approaches. This has made it difficult to distinguish which studies actually draw on the tradition of audience research (see also Mathieu et al., this issue). We have therefore opted to remain open in our collective understanding of online audience research and included all kinds of empirical research, as long as these audiences were investigated in relation to a text, media or technology.

We have centred our literature search around the collection of journal articles, although we have included other publishing formats as well. We opted for this preference in order to seek out empirical contributions over theoretical works that may aim to influence the field, but which do not always have an impact on conducting empirical research. Our collection of targeted publications comprised of, on the one hand, thirteen international
leading journals in media and communication research and, on the other hand, fifteen European journals with a mainly national scope. Table 1 synthesizes the primary outlets included in our bibliographical search:

Table 1: Media and communication journals included

<table>
<thead>
<tr>
<th>International journals</th>
<th>National journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Research</td>
<td>Comunicação e Sociedade (POR)</td>
</tr>
<tr>
<td>Communication Review</td>
<td>Comunicación y Sociedad (ESP)</td>
</tr>
<tr>
<td>Convergence</td>
<td>Comunicar (ESP)</td>
</tr>
<tr>
<td>European Journal of Communication</td>
<td>Comunicazioni Sociali (ITA)</td>
</tr>
<tr>
<td>European Journal of Cultural Studies</td>
<td>Estudos em Comunicação (POR)</td>
</tr>
<tr>
<td>Interactions: Studies in Communication and Culture</td>
<td>Media &amp; Jornalismo (POR)</td>
</tr>
<tr>
<td>International Journal of Child-Computer Interaction</td>
<td>Media &amp; Viestintä (FIN)</td>
</tr>
<tr>
<td>International Journal of Communication</td>
<td>Media Perspektiven (GER)</td>
</tr>
<tr>
<td>Journal of Children and Media</td>
<td>Medien Kultur (GER)</td>
</tr>
<tr>
<td>Journal of media practice</td>
<td>Obs* (POR)</td>
</tr>
<tr>
<td>Media, Culture and Society</td>
<td>Rassegna Italiana di Sociologia (ITA)</td>
</tr>
<tr>
<td>New Media &amp; Society</td>
<td>Réseaux (FRA)</td>
</tr>
<tr>
<td>Participations</td>
<td>Sociologica (ITA)</td>
</tr>
<tr>
<td></td>
<td>Studi Culturali (ITA)</td>
</tr>
<tr>
<td></td>
<td>ZER (ESP)</td>
</tr>
</tbody>
</table>

When selecting these journals, we took advantage of CEDAR’s geographical and cultural diversity in order to also select other-than-English-language publications, which may play a more marginal role in international discussions. The criteria used to select the international journals were mainly driven by the outcomes of in-group queries and led to identifying those outlets in which audience research was published most often. Additionally, we used libraries and online search engines, like Google Scholar, Web of Science, JStor, Ebsco, SSCI, Sage as well as Taylor & Francis search indices in order to enrich our database with complementary literature. We also added literature after consultation of the retained material, as well as on the basis of our own expertise. When justified, we have also included references outside the chronological scope of our review.

Our search outcomes were organized in a common Zotero library and filtered afterwards according to the specific issues addressed by this paper (see ‘the five lenses’ below). We collected over 850 items, out of which we retained 390 specifically relating to online audience research. We are aware that this represents a fraction of all the literature available, but we believe we collected, discussed and reviewed a substantial corpus that reflects the depth and variety of online audience research.
The five lenses of the literature review

After an initial discussion of the first literature corpus, we started focusing specifically on the following issues, as they emerged as central and because they provided some concrete entry points into online audience research. This methodology allowed us to look at the basic repertoire of our common research field, in that these issues were intended to represent the main questions asked by audience research:

1) The conceptualization of online audiences as *produsers* - *Who* are the online audiences?
This topic has generated a massive amount of research along various interests. Hence, it represents well, both in terms of its quality and diversity, some of the most interesting shifts in audience research over the past ten years.

2) The difficulties to keep track of online audiences due to the convergence and digitalization of media - *Where* and *when* are online audiences to be found?
We assume that media convergence and digitalization have influenced profoundly the methodological direction of online audience research, which follows from the ways audiences’ uses of media have changed as a response to a changing media landscape.

3) The possible reunification of reception and use in online audience practices - *What* do audiences do online?
A noticeable phenomenon brought by online media concerns the reconnection of uses and interpretations as part of the online practices of audiences (Livingstone, 2004). This presents the opportunity to investigate reception in conjunction with media uses, bridging a gap brought by broadcast-era research that kept them separate.

4) The ethics of online audience research - *How* should we study online audiences?
New and long-established methods of online audience research call for a re-negotiation of research ethics. We looked at how this challenge was addressed by the literature.

5) The knowledge interests guiding the study of online audiences – *Why* should we study online audiences?
Habermas’ categorization of knowledge interests (1972) – practical (hermeneutical), emancipatory (critical) and technical (predictive) – helped us capture some essential differences in the aims of online audience research.

The transition towards online audience research: an overview

In a research landscape in which it becomes increasingly difficult to distinguish audience research from other fields of research, it is crucial to be explicit in the standards and premises orienting research. But our first observation is that the specific challenges faced by
online audience research seem to be primarily discussed in theoretical texts, and are not often spelled out or reflected upon in empirical research. Most empirical studies do not state explicitly or discuss why they choose particular methods to study a particular question, let alone reflect on the results in relation to the chosen method. For example, although the reunification of interpretations and uses appears as a promising opportunity, which addresses issues that have plagued qualitative audience research since its beginnings and was emphasized in an influential article (Livingstone, 2004), its methodological opportunities are neither thoroughly discussed in the literature nor really embraced by empirical studies (as in for example Yoo, 2011).

**Diversity of knowledge interests**

While we came to observe this gap, we also identified a broad heterogeneity of interests animating research about audiences on the Internet and new media. The vast majority of academic online audience research falls into the category of hermeneutical knowledge interest, as it aims to interpret and understand audience practices with new media. Arguably, this knowledge interest can be said to have motivated much of the transition towards online audience research, often driven by the motor of technological novelty. The hermeneutical interest is served by a broad variety of methodologies, both quantitative (surveys and content analyses; Gonzalo et al., 2014; Tabernero et al., 2008) and qualitative (interviews and ethnographies, Mascheroni, 2013). In comparison, the critical and technical knowledge interests have animated a relatively small portion of the research. Also a clear divide between qualitative and quantitative research can be identified, recalling a traditional divide in audience research (Barker, 2006).

Discussions relating to the technical knowledge interest can be related to the rise of computational methods, ergo automated analytical tools and big data, both within academic audience research (Golbeck & Hansen, 2014; Savage et al., 2014) and in the media industry (Webster, 2014; Napoli, 2011; Minelli et al., 2012). Recently, Anderson (2011) claimed that the knowledge interest of online audience research is transiting from participatory (hermeneutics) to algorithmic (technical). Furthermore, Lee et al. (2014) suggest that ‘sophisticated tools’ enabling the tracking of audiences are the reason why media companies are interested in their audiences in the first place. Rather surprisingly, given the critical perspective that has informed much of (new) qualitative audience research since its inception, the critical knowledge interest orient only a minority of research. Few empirical studies amongst those reviewed aim explicitly at emancipation and at challenging dominant structures (but see Matthews & Sunderland, 2013; Rybas & Gajjala, 2007; Taylor et al., 2014; Lahey, 2014; Kuehn, 2013), and even fewer let the critical knowledge interest shape their methodology, for example by employing participatory methodologies (Franquet et al., 2011). An explanation could be that the critical knowledge interest is often informed and accompanied by substantial theoretical work, hence escaping the scope of this literature review.
Reliance on and challenge to existing research procedures

More often than not, the starting point to do research online is pre-online research (Muñoz, 2007). Long-established research methods, developed prior to the advent of online media spaces, i.e. survey, interview, ethnography, are still the main toolkit for scholars. Moreover, traditional methodological standards are maintained, even when the specificities of online spaces are recognized as the main focus of analysis, such as in studies that rely surveys to study online interactivity (Reinhard, 2011a; Chung & Yoo, 2008; see also Weinstein, 2014 or Heikkilä & Ahva, 2015). Studies of produsage, participation and fan or amateur production are often carried out with the help of relatively conventional methods of research, such as content and discourse analyses (Wagener, 2014) or surveys, often in an attempt to establish causal relation between online media use or access and democratic participation (Ognyanova, 2013; Östman, 2012; Mossberger et al., 2012). Another example can be found in how research ethics are implemented and reflected. Here ‘the human subject model’ (Ampofo, 2011: 29) of pre-online research has served as a reference for developing ethical standards for online research. Well-established methodologies are also being applied to online environments, giving rise to new approaches to the study of audiences, such as the use of conversational analysis (Steensen, 2013) or heuristic model of audience inclusion in journalism (Loosen & Schmidt, 2012).

Existing research procedures are also challenged by online audience practices, which are argued to be different from their offline counterparts. New gratifications are emerging from the affordances of new media, challenging existing research procedures (Sundar & Limperos, 2013). Audiences have developed an interest towards personalised web content (Golbeck and Hansen, 2014), which brings about concerns regarding the fragmentation of the audience and the formation of media enclaves (Beam, 2014). Exposure has changed from non-selective to selective and more conscious (Würfel, 2014; Graf et al., 2008) and, accordingly, audience research seems to have moved from a text-centric to a context-centric or practice approach¹ (Couldry, 2011). The latter can be substantiated by the number of ethnographic studies in the field or by the choice of specific practices that are considered worthwhile studying, for example activism.

Orientation towards the online traces left by the audience

The most revolutionary aspect of online audience research, compared to broadcast-era media research, is the analysis of online traces left by audiences throughout their uses of technologies. This methodological opportunity has been interpreted very differently by various strands of research, which created tensions across viewpoints, including crucial ethical considerations. For example, this opportunity has fuelled the use of unobtrusive methods of automated tracking (Kahn et al., 2014; Hight, 2015), measuring exposure (Graf et al., 2008), retrieving usage (Golbeck and Hansen, 2014), or testing media selection criteria (Beam, 2014). While these are perceived as an improvement in validity compared to traditional obtrusive methods, and may be said to resolve the ‘observer’s paradox’ (Labov,
1972) that has plagued pre-online methodologies (Schrøder et al., 2003), they also challenge existing ethical research practices (which consider obtrusiveness not as a problem, but as a way to obtain consent and establish rapport with research subjects) and may be said to contribute to the increased surveillance of citizens. It should also be mentioned that these traces have often been attached to the signifier ‘user’ in the context of user-generated content (UGC), produsage or participation2, a signifier that has largely gone unquestioned and uncriticised in empirical research (but see Flanagan, Hocevar & Samahito, 2014; Dynel, 2014; Livingstone, 2013; Meyen et al., 2010).

With regard to the different ways traces left by audiences are processed, research can be divided into two main camps, which reflect traditional divides between qualitative and quantitative traditions of audience research. On the one hand, a strand of research inspired by online ethnography considers the qualitative and contextual dimension of these traces. This is the case with studies that contextualize online practices, such as produsage or participation, in the offline world of participants, relying on traditional methods such as interview or observation (Goode, 2010; Binark et al., 2009; Callén et al. 2007). On the other hand, other studies follow the model offered by big data, applying numbers and not engaging in a contextualization of these traces. Within this trend can be ranged studies that compare news stories displayed by journalists and chosen by readers (Boczkowski & Peer, 2011; Boczkowski et al., 2011), studies that detect audience preferences or political homophily via Twitter data (Colleoni et al., 2014), or that identify the relation between audience clicks and news placement on websites (Lee et al., 2014).

The possibilities provided by the traceable presence of online audience have led to a great deal of methodological innovation. Complex multimedia audience practices have begun to be analysed by new computational methods now applied to the study of audience such as network analysis (Scott, 2011; Yuan & Ksiazek, 2011; Webster & Ksiazek, 2012), analysis of user-defined repertories (Taneja, Webster & Malthouse, 2012), and online tracking systems (Vicente-Mariño, 2013). Also worth mentioning is the recent development of mobile methods, which remain rare (Berg & Düvel, 2012; Fay, 2007). The possibilities of quantification and visualization of digital traces have produced interesting, but not always insightful innovations, which at times testify more to the availability of the data and computational power (retrieving a large amount of tweets), than on thoughtful ontologies or epistemologies of research. But this is not purely a potential threat to quantitative research, as online ethnography is also trying to make sense of the wide availability of ‘found data’, which according to Hine (2011) can be conceptualised as ‘multi-sited exploration’ (see Mascheroni et al. 2008 for an empirical example) and ‘itinerancy’. Hine’s work provides an example of how online ethnography has reinterpreted its canons in the face of Internet (see also Hine, 2015, 2000; Bengtsson, 2014; Kozinets, 2010).

In terms of adaptation of traditional methods, the growing number of content analyses is worth mentioning, both of the qualitative (Hughey & Daniels, 2013; Silva, 2013) and quantitative (Milioni et al. 2012; Tenenboim & Cohen, 2015) kinds (which dichotomy has become more like a continuum with many middle-range possibilities, as for instance...
Marichal, 2013), and their attempt to substitute traditional reception analyses, such as in Tenenboim & Cohen (2015), or studies of media uses, as in Milioni et al. (2012) or produsage, as in Wagener (2014). The latter study is particularly telling in its application of content analysis to explore reader’s comments of news as a way to assess audience participation to the public sphere, as it consequently examines uses (or functions) via reception.

**Four main methodological trends in the transition towards online audience research**

In the following, we offer a synthesis in the form of four main interconnected, but at times also contradictory, trends that we identified as dominating the landscape of online audience research and served to organize research in its transition from pre-online to online research between 2005 and 2014. These four methodological trends are: 1) the expansion of online ethnography and the continued importance of contextualization, 2) the influence of big data and a disproportionate emphasis on uses compared to questions of reception, 3) the reliance on mixed methods and the convergence of different research rationales previously kept separated, and 4) the ambiguous public nature of online data and the crucial ethical considerations for the conduct of online audience research.

1) **The expansion of ethnography and the need for contextualisation**

Multiple labels denoting ethnographically inspired online research can be found, such as online ethnography, cyberethnography, social media ethnography, digital ethnography, ethnography in virtual space, internet ethnography, netnography, etc. (Larsen & Glud, 2013; Postill & Pink, 2012), but all these methodologies converge towards the study of audience online practices, often oriented towards productive or performative uses of online media or produsage (Flanagin, Hocevar & Samahito, 2014; Dynel, 2014; Graham & Hajru, 2011). The methodology of online ethnography, with its versatility and its emphasis on contextualization, has witnessed a constant development since the advent of internet (see Hine, 2015, 2000) – to the point where an argument could be made that the hermeneutical knowledge interest has taken an ethnographic turn in online audience research.

The expansion of online ethnography has permeated many corners of online audience research, such as the study of online communities and processes of identity formation (Mihailidis, 2014; Willem et al., 2010), UGC, participation and produsage (Macek, 2013); fandom and amateur production (Evans & Stasi, 2014; Freund & Fielding, 2013; Gray & Mittell, 2007; Prieto Blanco, Schuppert and Lange, 2015), activism and social movements (Alexander & Aouragh, 2014; Camerini & Diviani, 2012), and gaming (Isabella, 2007; Gurney & Payne, 2014), to name but a few.

The development of online ethnography is not specific to online audience research, and its import does reveal tensions between offline and online contextualization (Dirksen, Huizing & Smit, 2010) when applied to audience research. Online ethnographic studies are
often conducted on the premise that online contexts are places where cultures are formed and practiced, as much as they do in offline contexts. In audience research, context is not a site of research, but rather a procedure for the proper interpretation of data regarding media use and content.

Accordingly, online ethnography is interested in the situated contexts formed online, especially as these are seen to (re)shape time, space, interaction, identity or other contextual dimensions (Marwick & Boyd, 2011; Takahashi, 2010; Willem et al., 2010; Verboord & Van Luijt, 2009, Van den Broeck et al., 2007). Illustrative examples can be found in Kanayama (2003), who investigates how elderly people interact and construct relationships online, based on the presumption that offline contexts of interaction are reorganized online, or in Meißner (2014), who argues that online contexts value and develop opinion leaders in ways that were not possible offline, or similarly in relation to fandom, which is said to take new meaning in the age of internet (Lee, 2011a).

In audience research, contextualization often expresses the need to ‘understand digital media in context’ (Burgess et al., 2013: 2, with the implication that digital media is not a context). Such offline contextualization often relies on the survey method in order to find out who are the respective people involved online, what is the nature of their engagement and what implications do online practices have on their offline life (Štětka & Šmahel, 2008; Bakardjieva & Smith, 2001). As such, audience research tends to regard context as an offline reality, in terms of socio-demographics (Corner, 1991) or life history (Radway, 1991/1984; Tufte, 2001). This contextualization has proven useful for media literacy (Davidson & Martellozzo, 2013; Sonck et al. 2012; see more generally the research project EU Kids Online), for understanding news consumption patterns (Waal & Schoenbach, 2010) or for promoting public pedagogy through connecting real life determinations with the cyberspace (Binark et al., 2009).

At times, the motivation for engaging in offline contextualization is grounded in the presumption that online media and technologies do not reshape, but prolong offline patterns (Dirksen et al., 2010; Kozinets, 2010). For example, Van Cauwenberge et al. (2010) see the same gratifications at play for the consumption of online news. Booth & Kelly (2013) find that many aspects of fandom remain unchanged by digital technologies, which instead helped with expanding the scale of these practices. Similarly, the ‘convergent culture’ identified by Jenkins (2008) seems to be the result of intense activity of specific categories of highly engaged and technology-savvy audiences, while the vast majority remains lurkers or recipients of digital media (Nielsen, 2006). These findings demonstrate how an important rationale of pre-online audience research needs to be maintained for the study of online media uses.

However, the crossover between offline and online contexts is not without problems. Research designs striving for offline contextualization often rely on obtrusive methods, such as interviews, diaries or surveys (Pedroni et al., 2014; Zhang & Lin, 2014; Zúñiga et al., 2013; Goode, 2010), and do not take advantage of the possibilities of scrutinizing online data. In fact, we observe a tension between old and new rationales of
conducting research concerning the possibility to engage with the characteristics of online data, such as its public nature, the anonymity of its participants, the presence of algorithms, or the capacity to engage with large volume of data. For example, the case of automated information technologies (AIT), which disguise themselves as humans and as such appear to act as conscious agents, brings challenges to conventional application of ethnography.

The question of ‘reduced social presence’ (Bengtsson, 2014: 865) seems to be a source of tension in conceptualizing online contexts (see Wojcieszak et al., 2009, for an illustration with the concept of ‘online deliberation’). For example, Morozov (2011) denigrates ‘slacktivism’ for its lack of resemblance with (real-life) activism (but see Karpf, 2010 for a counter-argument). Such criticism contributes to articulating an exaggerated dichotomy between online and offline spaces, which supports the creation of hierarchies of value between offline data (true, reliable, consequential, etc.) and online data (fake, unreliable, inconsequential, etc.), with the implication that online practices must be appreciated with reference to offline realities and not on their own terms. But as digital spaces are becoming less exotic, more mundane, and an integral part of everyday life, the distinction between online and offline becomes less relevant (Jensen, 2014; Fornäs, 2008).

Teli et al. (2007) argue that online ethnography is not always applied as holistically (taking account of all the relevant viewpoints and agents) as has been the case in offline ethnography. This, according to the authors, results from a false understanding of online spaces as environments detached from offline context. To counter this trend, recent research suggests a multi-sited and connective ethnography across online and offline practices (Larsen & Glud, 2013). This rationale has for instance been applied in Heikkilä & Ahva (2015), who acknowledge the influence of both online and offline social networks for the relevance of news.

2) The rise of big data and its emphasis on use
The rise of computational power, the proliferation of digital media and thereby of digital footprints, the increasing availability of analysis software, the coming of researchers formed in computer sciences into the scene of communication studies and of course the sheer amount of data available have seen the emergence of what is now known as big data. Although the reliance on computerized databases is not new, the emergence of big data creates a radical shift of paradigm in social research more generally (Conte et al., 2012), as it reframes key issues regarding the foundation of knowledge, the processes and techniques of research, the nature of information, and the classification of social reality (Lazer et al., 2009). Big data forces us to reconsider our theoretical foundations concerning social networks, communication, public opinion and social influence (boyd & Crawford, 2012).

This interdisciplinary field of inquiry at the crossroad between computer science and social science makes use of computational methods to analyze and model social phenomena related to audience research. Increasingly, the audience is instrumentalized and commodified as big data, as quantifiable and algorithmic, by media and other big corporations (Anderson, 2011; Bolin, 2014; Sponder, 2012; Napoli, 2011). Twitter and
Facebook have been and continue to be particularly well-studied by researchers interested in the networked relations of users and other aggregate phenomena of public opinion, expression and organization (Tinati et al., 2014; Bredl et al., 2014; Valenzuela et al., 2014; Moe, 2012). These social media are often analyzed by complex and automated algorithms, whose operational logics have become increasingly difficult to understand for the qualitatively-oriented audience researcher.

Big data is also symptomatic of the tendency to develop methodologies of research motivated by and aligned to technological developments. This can be witnessed in the many studies offering cloud-like visualisations of data, whose technological prowess seems at times to overshadow its relevance and use for research. Such ‘data-driven research’ (Burgess et al., 2013) has also become the flagship of a technical knowledge interest and has seen a renewed interest for questions of prediction that excite the imagination of governments and media corporations (Napoli, 2011). In fact, big data may be threatening the very raison d’être of audience research, which is to provide knowledge about contexts and interpretations (Zeller, 2015). These developments resuscitate old dichotomizing debates on empiricist epistemology, which presumes an intrinsic significance and accuracy of its found data, versus an interpretative stance that insists on the need to interpret constructed data (Markham, 2013; Anderson et al., 2009; boyd & Crawford, 2012; Manovich, 2012), as well as debates on the merits of quantitative scope versus qualitative depth (Mahrt & Scharkow, 2013; boyd & Crawford, 2012; Manovich, 2012).

It becomes evident that the rationale of big data privileges the study of media uses to the detriment of the study of reception or audience interpretations. Indeed, the separation between a technical and a hermeneutical-critical knowledge interest tends to reproduce a separation between the study of uses and interpretations. Of course, this development is not simply attributable to big data, but reflects also a repositioning of audience research towards the productive uses of media (Willem et al., 2010), to which also online ethnography has contributed, as well as a rise in methods, such as content analyses, that substitute reception analyses, on the merits that they are unobtrusive and more encompassing (Hunt & Koteyko, 2015; Hughey & Daniels, 2013). The quality and quantity of online data has broadened the extension of content analysis to the study of online communities, to include factors as varied as content structure, administrator activities, editing privileges, etc. (Roth et al., 2008).

The reliance on content analyses to study reception has seen a quantification of the study of meaning (Camerini & Diviani, 2012), apparent in the study of online news audience (Milioni et al., 2012; Tenenboim & Cohen, 2015), in network analyses (Yuan and Ksiazek, 2011; Webster & Ksiazek, 2012), and in big data research (Obolei et al., 2012). By relying on Twitter data to explore offline political processes, McKelvey et al. (2014) provide an illustration of how a question of research that could have been explored through a qualitative reception analysis is being reinterpreted in an online context tainted by quantity. Alongside, the study of reception has also been replaced by the study of interaction, as a
result of two-way and many-to-many communication flows (Lewandowska-Tomaszczyzk 2015, Williams 2009), also in qualitative contexts of research (Steensen, 2013).

However, content analyses can at worst produce misleading and biased findings on reception, not only by ignoring a large part of users who continue to play the role of recipients (Bechmann & Lomborg, 2013) but also by misinterpreting the public nature of traceable data, which cannot inform us about private or other contexts of interpretation beyond visible usage. Indeed, research developed on grounds of quantification does not contribute to illuminating aspects of reception acknowledged as relevant among audience researchers, such as motivation, comprehension, discrimination, position and implementation, to rely on the model proposed by Kim Schrøder (2000). We also know from past audience research that ‘content’ and ‘interpretation’ are not equivalent.

The application of these methods effectively turns the audience into a text, i.e. into an object of research that can be contrasted with the notion of audience as subject or agent widely acknowledged in traditional audience research (Sullivan, 2013). This objectification poses ethical questions to the study of UGC and user interactions, in particular when these studies are conducted without the explicit consent of their authors (boyd & Crawford, 2012).

Textual or computerized analyses of the audience bring obvious limitations for the capacity of research to contextualize data, interpretations and findings. As questions of sense-making and experience have been underprivileged in networked media analyses, the danger is to assume that networked media reception processes are similar to broadcast media reception processes (Mathieu, 2015). Hence, we see reception analysis to play a more prominent role in the transition of research towards online audiences. We see it as a challenge for research to be able to combine big data (volume and velocity) with ‘smart data’ (Caragliu et al., 2011) (veracity and value), the latter associated with a need for meaning and contextualization. Some new research has begun to answer the challenge: While Wilson & Dunn (2011), for instance, rely on a mixed method design based on self-report of media use associated with protest-related communication and Twitter data related to transnational audiences, Anderson et al. (2009) rely on ethno-mining, a combination of ethnography and big data.

3) The development of mixed methods and the convergence of research rationales
The application of mixed methods follows profound changes in the visibility, volatility and distribution of audiences across the media landscape (Perälä, 2014). As such, it does not specifically concern online audience practices, as most studies tracking cross-media uses do so across traditional and digital media (Schröder, 2015; Bjur et al., 2013). The reliance on mixed methods is not only the result of a more complex environment, but also the outcome of an epistemological motivation to transcend the qualitative-quantitative divide (Courtois et al., 2015; Dhoest, 2012; Mathieu, 2012; Nyiro, 2012; Schröder & Kobernagel, 2010).
What is called mixed methods tends to refer to different realities. For instance, qualitative and quantitative rationales of research are being combined in the Q-methodology (Davis & Michelle, 2011), a method that relies on qualitative card-sorting combined with factor analysis. Sometimes, the alliance takes place within the same method, as in a hybridisation of quantitative and qualitative content analysis (Lewis et al., 2013). In other instances, methodologies seem to have fused together as a result of online environments of research. This is the case in Lamerichs (2013) or Hine (2011), whose ethnographic approaches to online data tend to respond to the need of conducting traditional reception analysis.

While the notion of mixed methods refers to a specific and well-established research practice today, the trend englobes a much broader reality. Not only is the media environment converging, but recent work also adopts a converging view on a multitude of rationales, and these affect research methodologies. Convergence has proven to be a rich concept in online audience research, pointing to alliance or integration in term of context: between broadcast and network media (Jenkins, 2006), geographical contexts, consumption and production (Bruns & Jacobs, 2006), in terms of conceptualisation: the public and the private (Mascheroni et al. 2008), one-way and two-way flows of communication, in term of theorisation: political economy of media and actor-network theory (van Dijck, 2013) or fandom mixed with gaming (Wilson, 2011), or in term of media studies being integrated with other disciplines, i.e. big data.

One area of research that has seen innovative methodological exploration in response of a changing media landscape is the area of mobile methods (Taipale & Fortunati, 2014; Berg & Düvel, 2012; Goggin, 2011). Not only have these methods been developed to be able to tackle new qualities of media uses, such as ‘ubiquity, connectivity and convenience’ (Aguado & Martinez, 2009), but they also combine rationales of research that have proven difficult to reunite until now, concerning not simply quantitative scope with qualitative depth, but also a combined reliance on big data and contextualization to simultaneously study aspects of uses and interpretations (for a discussion, see Ormen & Thorauge, 2015).

We see these rationales of convergence to be important in order to transcend the limitations of single methods that keep things separate in an increasingly complex and interdependent media landscape (Lee et al. 2014; Trilling et al. 2013; Mitchelstein et al. 2010). But a critical view can also be taken, questioning whether these different rationales really converge (for example, the qualitative and quantitative), whether they should converge (the public and the private), and whether other rationales should also be converging (as in the missed opportunity of reuniting uses and interpretations, or big data and contextualization). We can also discuss our capacities to face all this convergence, as big data may confound qualitative researchers and, vice-versa, ethnography disconcerts researchers versed in statistical analysis, especially as this capacity building rests upon the establishment of university disciplines and departments and the education provided to
current and future scholars in media and communication studies (boyd and Crawford, 2012; Manovich, 2012).

4) The ambiguous nature of online data and its ethical challenges

The literature review identified a grey area concerning the private versus public dimensions of data, which brings about uncertainties to traditional research procedures and which awaits a more decided characterization via empirical research. This can be seen in the way the public nature of meanings is not properly recognized and taken advantage of in research, but also in the ethical dilemmas that follow from the ambiguity in the ownership of data.

The novel and ambiguous nature of online data, in the form of traces left by media users, create ethical challenges for online audience research. Here, as for many other challenges posed by online data, the point of departure has been pre-online research. Social research in various disciplines has been informed by The Nuremberg Code (Ampofo, 2011), which defined ethical principles, in particular for natural-scientific, experimental research. These principles included the essentiality of voluntary consent and the participant’s liberty to bring the experiment to an end. Meanwhile, it remains questionable whether the underlying human subject model of the Nuremberg code can successfully be applied to offline and online social research with its specific challenges (Eynon et al., 2008).

Ethical guidelines for online research developed in 2002 by the Association of Internet Researchers (Ess & AoIR, 2002) drew on the human subjects model, a choice whose limitations for conducting online research were subsequently criticized (Eynon et al., 2008; Sveningsson, 2004; cf. Estalella & Ardèvol, 2007). In 2012, the AoIR issued a revised version of their ethical recommendations (Markham & Buchanan, 2012). The document implicitly answers to the critics’ comments by picking up on process-focused ethical propositions especially promoted in dialogical-situational ethics (Allen, 1996; cf. Estalella & Ardèvol, 2007) or in feminist situated ethics (Piper & Simons, 2011), such as ongoing and renegotiated consent (Miller & Bell, 2012/2002). It explicitly calls for a process approach to ethical decision taking, which acknowledges the ongoing character of ethical decision taking and the necessity to incessantly and reflexively re-evaluate measures taken. While the document unambiguously states that ‘the greater the vulnerability of the community / author / participant, the greater the obligation of the researcher to protect’ (Markham & Buchanan, 2012: 4), it clarifies that some ambiguities and tensions within the field cannot be easily transgressed. This encompasses controversies regarding the concept of the ‘human subject’ in internet research, the blurriness of public and private spheres of data and action (cf. also Courtois & Mechant, 2014; Bredl et al. 2014), as well as the relationship between data/text and persons.

General challenges of online audience research that are mentioned in the literature concern: (1) soliciting participant consent, (2) appraising the trustworthiness of online audiences who use anonymized online identities, (3) drawing the line between what are
public and what are private spaces online, (4) less co-presence between researchers and participants so it becomes more difficult to judge whether participants come to harm, (5) securing data protection when storing and submitting data that are collected online and securing anonymity when data are published, and (6) insufficient legal boundaries that lag behind technological developments (Ampofo, 2011; Bredl et al., 2014; Eynon et al., 2008; Heise & Schmidt, 2014; Heise, 2015; Teli et al., 2007). Furthermore, especially in relation to social media data, produsage and UGC, it is (7) difficult to clarify data ownership and the extent to which researchers may ‘aggregate social media for their own purposes’ (Ampofo, 2011: 31).

Only a few publications explicitly address these ethical challenges and propose solutions. We also notice different ways of tackling these challenges. One way is to adapt traditional audience research methods and offline procedures for obtaining informed consent (Heise & Schmidt, 2014; Hookway, 2008; Mathieu & Brites, 2015). Another is to include anthropologically-inspired methodologies such as participatory and performative approaches to online ethnography (Hine, 2000, 2011; Larsen & Glud, 2013; Pink et al., 2015; special issue on virtual ethnography: Domínguez Figaredo et al., 2007).

Traditional techniques can be adapted and hauled out of their traditional limits to better understand internet environments, such as proposed by Hookway (2008). In the case of UGC such as weblogs, Hookway (2008) also reminds us that researchers will always have to consider, as in traditional interviews and focus groups, reciprocal honesty, trust and the willingness to participate. Even if the blog is public material, questions of ethics arise when using these without expressed consent. For social media research in general, Heise & Schmidt (2014) propose a catalogue of questions to decide whether it is necessary to inform the subjects that are analysed and to obtain their consent. These questions include, among others, which degree of privacy can be assumed and whether the knowledge interest is embedded in the media artefacts or in the practices of individual subjects. However, even when participants give informed consent, we cannot assume that they realise the extent of personal information that they reveal to researchers online (Freund & Fielding, 2013).

An example of a field with manifold discussions and solutions for ethical issues is online ethnography. The unavoidable involvement of the researcher’s everyday life experience with both text and context of online practices is widely acknowledged in ethnography. This creates new possibilities for proactively tackling ethical challenges, among others by engaging alternative ontologies of researcher and researched (Chimirri, 2013). Estalella & Ardèvol (2007), for instance, propose to focus on how online–offline practices are mutually established by researcher and researched. Jones (2007) argues for a performativ social science that takes its point of departure in auto-ethnographic accounts of online practices. Callén et al. (2007) reformulate the notion of mutuality as solidarity and present a techno-activist ethnography, in which researcher and researched pursue a common political aim together, both with their research and beyond. Freund and Fielding (2013) equally suggest building a strong rapport and solid relationship with participants (case of online/offline fandom). In fact, Berg and Düvel (2012) report higher return rates of
online media diaries when they previously established an offline bond with participants in interviews. Another issue that arises in the literature is whether unobtrusive observation is the appropriate way of gaining information, for instance whether it hinders researcher from a deep immersion and access to hidden areas in online forums (Greschke, 2007; Hine, 2011).

**Recommendations**

While research seems to be stimulated by technological novelty, it also continues to apply previous rationales, however not always aligned to current realities. We witness islands of research and a lack of overall consensus and approach over the challenges facing the study of online audiences. In spite of a massive research activity, there remain gaps and underprivileged areas that call for a re-prioritization of research.

The current state of methodological work in online audience research reflects some tensions and dilemmas, which, without being novel, reflect the challenges facing audience research today. There are continuous, perhaps increasing gaps between critical and technical knowledge interests, which often reflect a dichotomy between novel ways of scrutinizing the digital footprints of audiences and the need for contextualization. This dichotomy sometimes results in other unintended tensions between offline and online research or between a separation of the study of uses and interpretations. It is as if research has to choose between unobtrusive methods or contextualization, or between traditional methods or big data.

We wish to seize the opportunity of having conducted this literature review to reflect on the direction that this transition can take and offer recommendations for future research that follow two trains of thought: more convergence and more ethics. We believe that ethical considerations can provide an agenda for a better integration of different aspects of online audience research, concerning for example the integration of qualitative and quantitative research, uses and interpretations, conventional research procedures combined with the reliance on possibilities provided by technological affordances. We insist that, if these methodological points were to have an impact, it would be essential for empirical research to provide more space for methodological reflections.

Convergence seems to characterize much of the methodological innovations found in online audience research and we see this as a trend that should be consolidated in future research. We encourage exploring and reflecting on other possible convergences that have until now gone unnoticed or marginalised. We see rationales of convergence to be beneficial, especially as they contribute to address complexities in researching online audiences. Specifically, we recommend 1) more convergence between stakeholders and knowledge interests, 2) the development of mixed methods that integrate different research rationales and 3) a more reflexive application of methodology with the aim of providing advances to our understanding of online audiences.

1) We recommend more interactions with stakeholders and between the different knowledge interests motivating research. The industry is often in the possession of large
quantities of data, struggling to make sense of it. Academics can provide missing interpretations grounded in systematic and elaborated methodologies that maintain and further develop high ethical standards. Moreover, the divide that has arisen between the technical and emancipatory knowledge interests keeps important methodological rationales separate, for instance between the study of numbers and the study of language, or between uses and contexts.

2) We recommend the development of mixed methods that meet the requirements of increasingly complex research rationales of convergence, regarding the integration between the study of uses and interpretations, between quantification and contextualization or between unobtrusiveness and contextualization. This integration is especially crucial if academic audience research is to take advantage of the quantitative turn provided by big data.

3) We recommend more explicit attempts to transcend the opposition between online and offline contexts. While audience research should maintain its focus on adequate contextualization of numbers and words that stand for evidences of online practices, it is important that this contextualization does not overly emphasise a dichotomy between offline and online practices, as it is becoming clearer that the relationship between the two is more fluid, ordinary and widespread. Convergence can help transcend these dichotomies, but if this is to be fruitful, it needs to lead to theoretical advances, rather than being suggested on suspicious grounds of increased validity (Barker & Mathijs, 2012). For example, it would be desirable that qualitative and quantitative powers of explanation be brought together, if they could contribute to the reunification of interpretations and uses. Whether or not methodological considerations are the result of rationales of convergence, we find it important to reflect more thoroughly on the use of different methods in the theoretical context of online audience research. Here, it could be useful to reinterpret traditional understandings of audience research in the context of online, digital, interactive, or networked environments, such as the text-audience metaphor (see Mathieu, 2015), in order to orient the use of unobtrusive methods, such as content analysis, or obtrusive ones, such as the interview, to be better aligned to the current environment.

Ethical questions remain under-prioritised, while they should play an important role in providing directions for future research. This may imply that academic research is to explicitly distance itself from the surveillance of citizens, and instead provides guidelines for the ways the traceability of data left by audiences becomes a positive and emancipatory force that respects the integrity of those involved in the consumption and production of media artefacts.

We issue three main propositions for the ethical conduct of online audience research: 1) a new discussion of guidelines in the scientific world and outside of it, 2) the call for more transparency as well as 3) a reflexive role for the researcher and her relation with the participants.
1) Several researchers agree that online audience research needs a further adaption of ethical guidelines to current technologies, media and situations (Eynon et al., 2008; Heise, 2015; Larsen & Glud, 2013). In addition to a discussion within academia, Heise (2015) suggests that researchers should also engage in societal discussions on online ethics, for example concerning big data. Outside of the academic world, ethics and online privacy issues are considered keystones for an inclusive knowledge society (www.unesco.org).

2) Online audience research would benefit from a more transparent discussion of online ethics (Eynon et al., 2008). Some of the publications reviewed for this article were entirely lacking a discussion of online ethics, while others dealt with it only as a footnote. More transparency in ethics procedures would ensure that researchers can learn from each other and be held accountable for their doings inside and outside the academic field (ibid., Heise, 2015).

3) Ethical guidelines should be complemented with an ongoing and situated negotiation of ethics with the research participants (Piper & Simons, 2011), for instance in terms of an ongoing negotiation of informed consent (Miller & Bell, 2012). Participants are thus understood as co-researchers who have and take a stake in explicitly influencing the research process and its ethics (cf. Teli et al., 2007; Chimirri, 2013).

Biographical notes:

David Mathieu is assistant professor at the Department of Communication and Arts at Roskilde University, Denmark. His current work focuses on audience and reception research, social media, research methodologies and social change communication, with an emphasis on the changing nature of audiencing in the age of social media. Contact: mathieu@ruc.dk.

Maria José Brites is assistant professor at the Lusophone University of Porto (ULP) and postdoctoral researcher at the Communication Sciences at Communication and Society Research Centre (CECS/UMinho). Her research interests include issues such as Youth, journalism and participation, Audience studies as well as News and civic literacy. She started the blog ANLiteMedia as a way to track these issues. Contact: britesmariajose@gmail.com.

Miguel Vicente-Mariño is an associate professor at Facultad de Ciencias Sociales, Jurídicas y de la Comunicación, Universidad de Valladolid (UVa), Spain, and Head of Department of Sociology and Social Work. His research interests are oriented towards audience and reception studies, research methods applied to Communication Research, environmental communication and public opinion inquiry. Contact: miguelm@gmail.com.

Inês Amaral is an assistant professor at Autonomous University of Lisbon and Instituto Superior Miguel Torga, and a researcher in Communication Sciences at Communication and
Society Research Centre- (University of Minho). Her research interests include cyberculture, new media, social media, social networks, social movements, lusophony, media literacy, digital inclusion and active ageing. Contact: inesamaral@gmail.com.

Niklas A. Chimirri is assistant professor at Roskilde University’s Department of People & Technology. His research centers on psychological, sociomaterial studies of technology, communication and design in everyday life. Currently he investigates young children’s meaning-making through digital technology and develops participatory approaches to qualitative inquiry across age groups. Contact: chimirri@ruc.dk.

Juliane Finger is research assistant at the University of Hamburg, Germany. Her main research interests are subjective perceptions of media effects, and media and collective memory. Contact: juliane.finger@uni-hamburg.de.

Bojana Romic is an independent researcher and video artist based in Copenhagen. She is a member of the Danish Council for Ethnic Minorities (REM). Her research interests include: new media audiences, reception studies, digital literacy, e-learning and aesthetics of technology. Contact: info@bojanaromic.com.

Minna Saariketo is doctoral researcher in the School of Communication, Media and Theatre at the University of Tampere, Finland. Her research concerns agency and algorithmic power in technologically mediated society with a special interest in developing the approach of critical technology education. Contact: Minna.Saariketo@staff.uta.fi.

Marisa Torres da Silva is assistant professor at Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa (FCSH/NOVA), Portugal. Her research interests include the relationship between the media and democracy, as well as the forms by which citizens can participate in the press and in the online environment. Contact: marisatorressilva@hotmail.com.

Liliana Pacheco is currently Ph.D. student at the Centre for Research and Studies in Sociology, Lisbon University Institute. Her research interests are related to generations, journalism, new media and culture. Contact: lilianaribeirapacheco@gmail.com

References cited:


Graham, T., and A. Hajru, ‘Reality TV as a Trigger of Everyday Political Talk in the Net-Based Public Sphere’, European Journal of Communication, 26, 2011, pp. 18–32.


Lazer, David et al., ‘Social Science: Computational Social Science.’ Science, 323 (5915) 2009, pp. 721–723.


Mathieu, David, ‘The Continued Relevance of Reception Analysis in the Age of Social Media’, *Tripodos*, 0, 2015, pp. 13–34.


Torres Da Silva, Marisa, ‘Online Forums, Audience Participation and Modes of Political Discussion: Readers’ Comments on the Brazilian Presidential Election as a Case Study’, Communication & Society / Comunicación Y Sociedad, 26, 2013, pp. 175–93.


Notes:

1. As Couldry explains, a practice approach demands ‘taking as our starting point what people are doing with, or in relation to, media’ (2011, p. 217), rather than seeing them automatically as an audience to a medium.

2. The interest in UGC, produsage and participation has led to a massive amount of research in contexts as varied as online media activism (Camerini & Diviani, 2012; Alexander & Aouragh, 2014; Millioni, 2012; Mascheroni, 2013; Kotilainen & Rantala, 2009), social media (Mihailidis, 2014), online journalism (Singer, 2014; Lee et al., 2014; Beam, 2014; Lima & Reis, 2012; García-Avillés, 2010; Franquet et al., 2011; Carpentier, 2011; Mitchelstein, 2011; Boczkowski et al., 2011, Boczkowski & Peer, 2011; Amossy, 2009; Paulussen et al., 2007; Hughey & Daniels, 2013; Canavilhas & Rodrigues, 2013; Loosen & Schmidt, 2012), gaming (Dena, 2008; Isabella, 2007; Gurney & Payne, 2014),
fandom (Whiteman, 2009; Dori-Nacohen, 2013; Lee, 2011a; Lee, 2011b) and amateur production (Jenkins, 2008) in an attempt to grasp the varied forms of engagement of audiences.