GovReg

WORKING PAPER SERIES

Does Quality Really Matter? Exploring Data Quality Assessment in the French Online Advertising Market.

Théophile Mégali

2021

RESEARCH

Dauphine | PSL

CHAIRE GOUVERNANCE ET RÉGULATION

Does Quality Really Matter? Exploring Data Quality Assessment in the French Online Advertising Market.

Théophile Megali, Ph.D.

Paris-Dauphine University_PSL, Dauphine Research in Management (UMR CNRS 7088). Associate Researcher, Governance and Regulation Chair.

Martin Quinn, Ph.D.

Department, University, City, Country

Corresponding author: Martin Quinn, martinquinn@ucp.pt

Does Quality Really Matter? Exploring Data Quality Assessment in the French Online Advertising Market.

Data is a central element in digital marketing, but in recent years, doubts have been raised about the quality of data generally sold or used for targeting purposes, following cases of fraud and abuse. In this study, we sought to understand how marketers evaluate the quality of the data used for targeting purposes. This empirical study is based on interviews of 22 professionals in the French market. Among the results of this survey, it appears that "quality" is multiple and varies according to the nature and uses of the data. It also reveals that the logic of performance evaluation still largely prevails, but the trend towards internal and external evaluation of data collected by advertisers tends to develop a form of "data hygiene". Finally, data quality can become a differentiating factor between competitors, but the current state of evaluation tools benefits mainly to the dominant operators in the online advertising market, Google and Facebook.

Keywords: online advertising; data quality; quality assessment

Introduction

Is there a problem with the quality of data used by marketers? Since few years, several statements from the marketing industry raised concerns about the poor quality of data used in targeted advertising campaigns. Notably, reproaches are addressed about the recency of data sold, accused of being often “two or three days out of date” (Digiday Editors 2018), inaccurate (Downie 2018) or even missing audience (Olenski 2018). The result is counter-productive, as an off-target campaign can harm the consumer experience and misses the point for what data has been purchased for. A recent US marketing report estimated that 21 cents spent on every media dollar was wasted due to poor data quality (Forrester 2019).

In addition to this quality issue, fraudulent data collection (Joseph 2019) or incompliance to privacy regulations (Jaye 2019) from data vendors are also creating mistrust among industry actors. Therefore, marketers are overtly expressing concerns about the quality of purchased datasets since several years in the US (eMarketer 2016) as well as in the French advertising industry (Del Frate 2020).

In this context, firms can be enticed to ask for evidences of quality when they purchase datasets. Historically, the need for a "market feedback" provided by various sources of data (e.g. sales, surveys, panels), as well as the ability to target a narrower segment of the audience emerged around the early 1910s in the US. The main idea, for advertisers, was to maximize their audience per dollar spent, which contributed to the development of market research initiatives (Beniger 1986). In the digital era, tracking costs have decreased significantly, thus fostering targeted advertising capabilities (Goldfarb and Tucker 2019) with much more precision than in offline media (Goldfarb and Tucker 2011). An abundant literature has been dedicated, in the last decade, to the analysis of targeted advertising and data tracking, assessing its efficiency from an economic and marketing perspective (Farahat and Bailey 2012; Lambrecht and Tucker 2013) or describing its features from a rather critical standpoint (Ebeling 2016; Turow 2012).

In this article, we analyze how marketers assess the quality of data used for targeting purposes and how it affects market practices toward a shift in data usage. Based on a grounded theory approach, we conducted a set of 22 in-depth interviews with executives from the French online advertising market. We review existing solutions regarding their meaning for practitioners, depending on the various sources of data.

In France, two landmark industry organizations have recently issued solutions to assess the quality of datasets. On one side, Mediamétrie is a partly public company which runs as a monopoly the panel-based audience measurement for TV and Radio broadcasts in France since the 1980s. Médiamétrie launched "Data Checking" in 2018, offering to match any sociodemographic database with its panel in order to check its accuracy. On the other side, the CESP (Center for the Study of Advertising Supports) is a French non-profit trade association created by industry actors in the 1950s to assess the methodology of advertising measurements and research. Since 2018, the CESP is offering to audit the processes and methods of data platforms, focusing on the origin, the collection, the treatment and the currency of data. These two solutions are innovating in the assessment of data quality, which was exclusively based on ex-post measurement of campaigns successes until then.

Our results suggest an ambiguous effect on the market. Firstly, we point out that two main types of data quality assessment exist: one *ex post*, based on the analysis of the performance of the campaign, and another one *ex ante*, developed by Médiamétrie and CESP, based on the evaluation of the database itself. These solutions aim at solving a classical information asymmetry by providing useful signals for data buyers (Akerlof 1970). They are broadly based on labels and certifications issued by specialized organizations on the basis of audits and grading of databases.

Secondly, our results show that the definition of quality varies, according to the intrinsic characteristics of data and their usage by practitioners, as well as the necessity for advertisers to assess them. This implies that these solutions remain incomplete in guaranteeing the quality of datasets, for technical and economic reasons.

Finally, we underline that this heterogeneity in such ability to measure the quality of datasets creates discrepancies among data providers. This contributes to alter the advertising competitive landscape as well as reinforcing the market power of major players.

This article provides the first comprehensive analysis of data quality verification techniques as well as their impact on the market. We provide a large set of results related to the adoption of new solutions allowing market actors to verify the quality of data when running online advertising campaigns. These results can be generalizable to information asymmetry problems on the markets. In our context, we analyze how a drop in information asymmetry related to the quality of data transform the online advertising market. Intuition may foster us to consider these solutions as always beneficial for the market: we show that the reality is more ambiguous. Our results also bear strong managerial importance as datasets appear to be usual inputs for advertisers and agencies running online campaigns. By analogy with *supply chain* concerns, the quality management of such components in the provision of the final product becomes an issue (Crosby 1979). In the context of this very industry, the quality problem exposed by industry actors could call for a collective solution and a form of regulation. Trade associations, market intermediaries and firms are endorsing this cause and adopting individual strategies to determine the quality of data they produce, purchase and use in their campaigns.

This article is organized as follows: we introduce the literature related to our topic and present our methodological underpinnings and the dataset we used. Then, we formulate hypotheses about this topic and discuss them, in the light of managerial recommendations. Eventually, we describe the limitations of our work and open the debate on further research.

Theoretical Background

Defining quality and its dimensions

Defining what is ‘quality’ when referring to a product such as data in the context of advertising refers to an abundant literature in various fields. Forging a global definition of the concept appears to be vain and its complexity can only be reflected through a cumulative, multidisciplinary vision (Reeves and Bednar 1994). In a synthetic paper, Garvin (1984) described the five main approaches of product quality:

(1) a philosophical view, where quality is transcendent, absolute and universally recognizable through experience and yet undefinable precisely;
(2) an economic product-based view, where quality is a precise and measurable variable relative to a ranking of attributes possessed by the product;
(3) a marketing or economic user-based view, where quality is idiosyncratic and tied to the personal views, needs and preferences of the consumer;
(4) a manufacturing-based view, where quality depends on the requirements and specifications expected in the production process;
(5) an operation management value-based view, where quality is determined by a provided performance at an acceptable cost.

This reflects the wideness of definitions of quality, but underlines the fact that in most of these approaches, quality remains a relative notion regarding a product in competition context. Chamberlin (1953) underlined the importance of quality as a vector of differentiation among products of the same kind in the consumer's decision. The products' characteristics are thus at the core of consumers' choice rather than the goods themselves (Lancaster 1966) and consumers compare and rank products according to the presence of these characteristics.

It is significant to remind that data could be defined in many different ways (Zins 2007), but an official definition is provided by the ISO/IEC 2382-1 norm which states that data is "a reinterpretable representation of information in a formalized manner, suitable for communication, interpretation or processing", as quoted in the 2014 Communication from the European Commission “Towards a thriving data-driven economy”. Within the specific industry of online advertising, data have become crucial assets, as advertisers make several uses of them. More precisely, data could be used in three main ways for enhancing the performance of online advertising campaigns: for pricing, analytics and targeting purposes. Data-driven pricing refers to the data used in the context of dynamic pricing activities and media buying strategies (Goldfarb and Tucker 2008). “Analytics” refers to the tools of measurement helping advertisers and agencies to optimize the resource allocations following indicators. Eventually, “targeting” means the personalization of the advertising content according to the features of the web user’s profile, determined through the collection and processing of data. The data produced by the users about themselves and their preferences can be used by advertisers to tailor marketing strategies and push advertisements to a targeted audience (Kumar et Gupta 2016). We will focus on this last type in this article. Targeted advertising also relates to the growing use of automated processes of inventories allocation (i.e. programmatic selling), which accounted for 78.5% of digital ad spending in France in 2018 (eMarketer 2019). In 2018, $340 million have been spent in the collection and processing of data for programmatic advertising campaigns in France (OnAudience.com 2018).

Works in economic sociology contributed to enrich this vision, assuming that the product results from a never-ending process of transformations in the definition of its characteristics by the economic actors and consumers (Callon, Méadel, and Rabeharisoa 2002). Thus, the product’s quality is a temporary configuration, overflowing its strict material characteristics and encompassing additional elements such as the “reputation of the seller” (Chamberlin 1933). Quality is therefore a collective construction process, calling for intermediaries and devices such as classification systems or standards (Beckert et Musselin 2013), rather than a self-generated and free-standing element.

Yet, existing articles on the definition of data quality are, to the best of our knowledge, mainly positioned in the field of computer science or information systems and adopt a quality management perspective, referring for instance to applicable ISO norms (Merino et al. 2016). These works mainly present lists of quality features that are required for data in a "big data" context (Cai and Zhu 2015). Among others, we find dimensions such as accuracy (i.e. closeness of results of observations to the true values), currency (i.e. the way data is up-to-date), timeliness (i.e. the ability of data to be available when expected), correctness, consistency (i.e. the ability for a dataset to encompass different perspectives), integrity (i.e. the degree to which data remain unified and altered), validity (i.e. the way data is conform to the syntax of its definition) or security (Cai and Zhu 2015; Gao, Xie, and Tao 2016; Lakshen, Vraneš, and Janev 2016; Loshin 2010; Sebastian-Coleman 2012). Usability is also mentioned, referring to the notion of "fitness for use" of data by data consumers described by Wang and Strong (1996). There is no consensus on an exhaustive list of quality dimensions that can be expected from datasets.

Information provision and assessment of the quality

Considering the relative nature of quality for products on a market, economists have investigated since decades the alternatives to the price signal. As demonstrated by the seminal paper of Akerlof (1970), price is not sufficient in presence of information asymmetry, which causes market failures, hence the necessity for providing quality certifications to better inform the buying side. In the selling side, good data quality providers are supposed to be enticed to signal their good quality, especially as it costs less for them (Spence 1978). These economic theory underpinnings paved the way for further research on the ability for actors to coordinate and organize such quality signaling process.

Various types of solutions can emerge in order to provide quality signals, when the quality cannot be immediately assessed by the buyers in a specific market. A supply of information can be provided by the seller, through advertising for instance (Nelson 1974). As well, Rao and Monroe (1989) have shed light on the impact of price, brand and store names on buyer's perceptions on quality.

Another solution is to provide information through norms that indicates the quality level, based on a global definition. Such coordinated devices can consist in licensing requirement (Leland 1979) such as occupational licensing or certifications (Shapiro 1986), but also standard-setting which reduces consumer's cost of quality assessment (Jones and Hudson 1996). These norms can be produced by a variety of institutional solutions, from public ordering to private organizations, as described by (Fernández-Barcala, González-Díaz, and Raynaud 2014).

When the offered product is diverse in qualities, and especially in the case of *experience goods* - i.e. its qualities are difficult to observe in advance, but can be determined after consumption (Nelson 1970)– standardization can be more difficult to perform. As analyzed by Karpik (1989), the buyer can rely, in this situation, on judgment from other buyers backed by trust and interpersonal relations. This judgment can also be provided through measurement systems. An historical inquiry by Velkar (2012, 2014) on the 19th century British wheat markets documented the institutional change and centralization of a measurement system. This enabled a shift from inspection to grading, based on voluntary consensus among market actors. Measurement systems are thus central in the development of economies and the 'manufacturing of markets' (Brousseau and Glachant 2014).

These measurement systems are often operated by intermediaries or 'third-parties', like

Médiamétrie or Nielsen in the audience rating market (Bourdon and Méadel 2014). This intermediation reduces the cost of access to information on quality for buyers (Stigler 1961) and become crucial in the transaction process as it tends to guarantee the quality of the product: this intermediary position is strategic in the construction of social relations among market actors (Bessy et Chauvin 2013).

As pointed out above, such quality signal should be valued according to the incremental return of having more information. According to the literature, such increase in level of information shall benefit to the data buyers (i.e. advertisers) and to the vendors providing high quality data: data quality assessment solutions, when available should thus be desirable to market players.

Materials and Methods

A recent trend in literature has investigated the way data are valuated as assets (Beauvisage et Mellet 2020), in specific fields like health services (Ebeling 2016). In this article, we propose to analyze, from market actors' point of view, what dimensions of quality are relevant in the context of targeted advertising and how it is taken into account in the valorization of data. We also introduce a distinction and a comparison between *ex ante* and *ex post* assessment for data quality in the French online advertising market. This is made possible since two organizations, Médiametrie and the CESP (Centre d'Etude des Supports de Publicité), issued new solutions to assess data quality for advertising purpose in 2018. Our aim is to describe and analyze how online advertising actors assess the quality of data and how this assessment affects the market shape and competitive structure.

Solution	Timing	Action
Digital Ad Ratings (DAR) - Nielsen	ex-post	Determining on a daily basis to what extent the ad campaign has been delivered and seen by the targeted audience.
Validated Campaign Essentials (vCE) - ComScore	ex-post	Determining on a daily basis to what extent the ad campaign has been delivered and seen by the targeted audience.
Data Checking - Mediamétrie	ex-ante	Scoring the accuracy of a sociodemographic database, based on a matching between this very database and a reference database ("panel") built by Médiamétrie.
Audit des plateformes data - CESP	ex-ante	Auditing the processes and methods of data platforms, focusing on the origin, collection, the treatment and the currency of data.

Table 1. Outlines of the different solutions mentioned in the article

This article is based on a *grounded theory* approach. The Grounded Theory emerged in the late 1960s thanks to Glaser and Strauss (1967), inspired by the previous works of E.C. Hughes who stressed the need for respecting empirical reality in the construction of abstract concepts (Demaziere and Dubar 1997). This approach suggests that concepts should emerge from the data collected on the field, and evolve concomitantly with the discoveries provided by the fieldwork (Glaser and Strauss 1967; Strauss and Corbin 1994). Previous theories shall not be ignored in this process, but neither determining hypotheses that lead the research work (Suddaby 2006).

In our article, our intention is to understand how practitioners are taking quality into account regarding data. We rely on a set of in-depth interviews with executives of firms or trade associations of the online advertising industry to document this. Types of firms were carefully chosen in order to have a representative view of the market and thus properly build a theory, coherently with the notion of *theoretical sampling* (Glaser et Strauss 1967; Creswell 2007).

From this, we built a contribution that paradoxically shows that quality information is not always desired from practitioners, even when available. Yet, available quality assessment solutions contribute to shape and build markets for data for targeted advertising purposes. These results are partially reconsidering existing economic theories. Our study also helps to build theory on the way digital technologies are bringing changes into the advertising practice, especially regarding to the importance of performance. Such insights can be brought with the help of Grounded Theory methods, that appear to be particularly fitted for investigations into such field (Goulding 1998), as existing theories are not successful in explaining the observed phenomena.

The Grounded Theory approach requires rigor in the collection and use of data (Corbin and Strauss 1990). To produce our results, we relied on a set of 22 interviews that have been conducted between September 2018 and July 2019, lasting between 30 and 120 minutes approximately. This volume of interviews appears to be sufficient to saturate our categories with enough information, as recommended by Creswell (p.64) who suggest a range of 20 to 30 interviews. We, the two authors, conducted most of these interviews together.

We established relatively flexible interview guides, with three main phases: a first phase of presentation of the interviewee's position in the firm and previous career, a second phase of technical or precise explanation of its use of data and a last phase regarding a broader vision of the advertising and data markets. We usually distributed the questions and follow-ups beforehand, in order to get a fluent interview proceeding. We recorded almost all the interviews (21 out of 22), in agreement with our interviewees. Interviews have been carefully transcribed to help the process of coding and sent to the interviewees when asked.

Pseudonym	Gender	Title	Firm	Type of companies
I.1	F	Head of Data Analytics	Multinational	Assessment Provider
I.2	M	Data Project Manager	Multinational	Advertiser
I.3	M	Partner & Cofounder	National	Agency
I.4	M	Data Project Manager	National	Publisher
I.5	M	International Project Director	Multinational	Assessment Provider
I.6	M	Product Lead	National	Publisher
I.7	M	Founder and CEO	National	Consulting Firm
I.8	M	Scientific Committee Member	Multinational	Assessment Provider
I.9	F	CEO	Multinational	Ad Tech Intermediary
I.10	M	Head of Operations	Multinational	Ad Tech Intermediary
I.11	M	Trafic Manager	National	Agency
I.12	F	Sales Manager	Multinational	Ad Tech Intermediary
I.13	M	CEO & Co-founder	National	Assessment Provider
I.14	F	Head of Analytics	Multinational	Advertiser
I.15	M	Lawyer	National	Consulting Firm
I.16	M	CEO	National	Data Provider
I.17	M	Chief Strategist Officer	National	Data Provider
I.18	M	Product Manager	Multinational	Data Provider
I.19	M	Managing Director	Multinational	Data Provider
I.20	M	Co-founder	National	Data Provider
I.21	M	Marketing Analyst	Multinational	Data Provider
I.22	M	Sales Manager	Multinational	AdTech Intermediary

Table 2. Description of Interview Participants.

Consistently with the recommendations of Corbin and Strauss (1990), we coded the interviews in successive steps, first proceeding to a first coding with descriptive items (*open coding*), and then, bringing abstraction progressively by comparing codes and exploring causes, strategies, intervening conditions and consequences related to the core phenomenon we explore (*axial coding*). From this, we eventually attained an advanced coding phase (*selective coding*), bringing up propositions that we relate hereafter, contributing to identify and assemble elements of our theory (Creswell 2007).

Beyond these primary data, we also relied on a corpus of grey literature (i.e. market research, reports, white papers), specialized press articles and tweets, but also academic articles in computer science of information systems. We also attended several professional events organized by trade associations, consulting firms and online press websites in order to enrich our global knowledge about the online advertising industry. These secondary sources were not included in the coding, but helpful to triangulate our results (Miles et Huberman 2003).

Results

Types and uses of data in the online advertising context

Within the online advertising market, data is sold in various ways and valuated for different purposes. Data is increasingly perceived as an asset (Beauvisage and Mellet forthcoming) and valuated for its usage in the advertising campaigns by advertising executives. In this context, market actors sell data in various ways we propose to present, depending on its nature and its usage.

Depending on the source of the data	
First-party data	These data are collected by the advertiser from its website, its customer database or stored from its previous campaigns.
Second-party data	These data are the first-party data from another firm, but sold, exchanged or given directly to the advertiser, often in the context of a partnership (e.g. an energy provider can exchange information with a telecom company about customers that are about to move house).
Third-party data	These data have been collected by data providers (e.g. Experian, Axciom) and are usually bought by advertisers to enrich their database with additional information, or to enlarge or restrict their audience.
Depending on the data collection method	
Transactional-based	Customer's data have been collected by an online/offline shop in the context of a transaction. For instance, data collected by fidelity cards are of this type.
Declarative-based	Web user's data have been collected via declaration form, surveys. These data have been filled out by the web user himself/herself.
Tracking-based	Web users' data have been collected by tracking methods and tools like cookies or IDFA when browsing on publisher's website. Beyond analytics or user experience purpose, the tracking enable to collect behavioral data (e.g. visited pages, time spent, "likes" on social networks, searchs, browsing sequence). Tracking methods are different on web and mobile environments, but data can be merged or cross-referenced (*cross-device*).
Depending on the nature of the information conveyed by the data	
Sociodemographics	These data are mainly used by marketers to build audience segments or *persona* (Hirth, 2017: 55). Contained information are relative to the personal, social and geographical situation of the individual. These data can be considered as permanent (e.g. age, gender) or rather temporary (e.g. job position, incomes, location, owner/tenant, single/married/…).
Preferences	These data focus on the interests of the web user: hobbies, cultural preferences, etc. Such data can be helpful for personalization or targeting purposes.
Purchase intentions	These data are more precise on the intention of the web user than preference information. Here, there is a variable degree of certainty that the web user has manifested an intention to buy the product. Such data can be acquired through transaction records, search queries, browsing, etc.

Table 3. A typology of data used in advertising campaigns

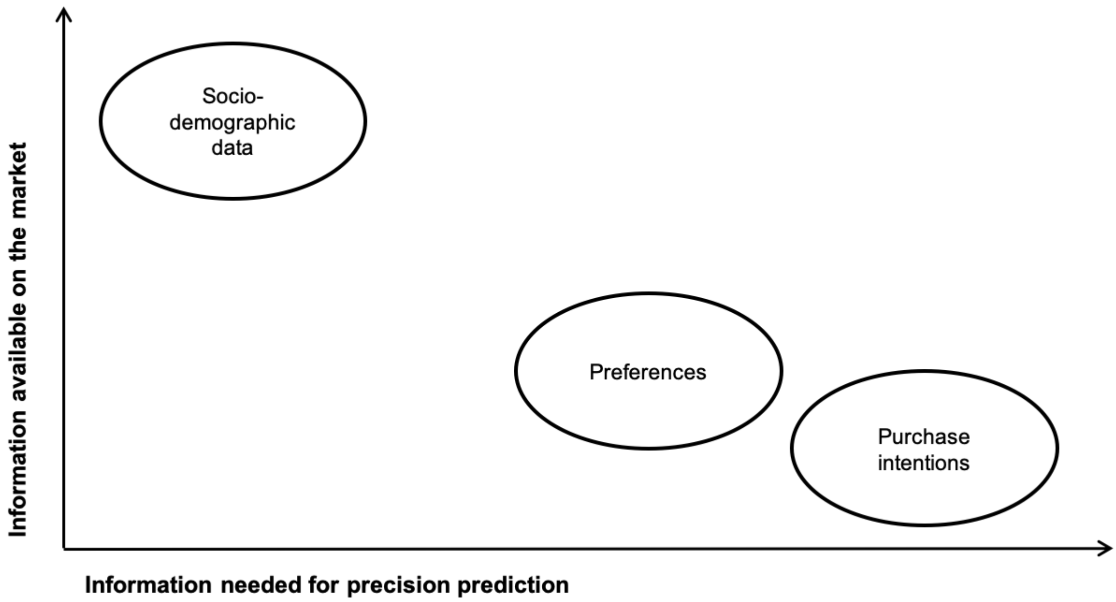
Beyond the notion of "big data" that conveys a quantitative vision of the data used in digital activities, a more precise debate occurs on a qualitative basis. Indeed, all data are not equal in the sense they do not convey the same informational charge, especially within an advertising context. We propose here a typology to classify the variety of data, depending on three factors taken into consideration for advertising campaigns.

From this typology, it appears that data can be valued differently depending on their *characteristics.* Advertisers use online information on users to predict customers' attributes. However, we find that some of these attributes may be very difficult to infer due to the lack of robust information on users. This may be the case of an attribute such as a "purchase intention that are mostly inferred from "tracking-based" data. Conversely, other attributes may be easier to predict for firms. This is the case of gender or age that are mostly inferred from "declarative-based" data, which therefore constitute more trustable source of information.

> "Probabilistic segments are not ready, so we do not sell them yet, but they will have a lower price than deterministic segments" (Interview I.6)

> "What we try to have […] in our CRM base (NB. Customer Relationship Management), is only deterministic" (Interview I.2)

These simple examples underline how predicting users' attributes directly depends on the type of information one is able to recover. However, the availability of information related to attributes may depend on those attributes. As just pictured in Figure 1, interviews revealed that preferences and purchase intentions are scarcer than sociodemographic data (age, gender, …).



For instance, if an clothes shop is willing to increase its online sales of a new backpack for bike commuters, it will rather target consumers that are potentially interested to buy such product. It will be easier for the shop owner to rely on data that reveals such preferences: data produced when web users that have, for instance, visited fashion pages on BikeMag.com, liked a Bike Commuters page on Facebook, watched bike fixing videos on Youtube or looked at the price of backpacks on Amazon.com. These data will provide more accurate and useful information to target more precisely potentially interested consumers than sociodemographic data (age, gender, etc.). On the contrary, if a car brand is willing to communicate on its values or brand image to a large audience, it will be more useful to rely on larger audience based on basic attributes (e.g. age 25-55, urban residents), thus using preferably sociodemographic data corresponding to the potential audience interested by the brand.

> “We focus on what we call personae, which means defining a number of characteristics of clients or prospects to address them the ad at the best moment […] like parents, mothers, return to school for instance, many small criteria we condense” (Interview I.2)

> "There are plenty of data that we can map to know, for instance, if the player goes to the web app […]. We can find touchpoints like this […] and so try to model things on its consumption during the past week, for instance" (Interview I.20)

It is significant to underline that data is also used reversely, to exclude a specific audience, when tailoring an audience. Such exclusion is made because a part of the audience is not necessarily concerned by the product (e.g. already client, not in the target) or because it has already been contacted. This step appears to be increasingly mandatory, as an erroneous or redundant targeting could have negative effects on web users.

> "Instead of reasoning about who I want to target and ask whether it is a good target or not, I rather think about who I absolutely want to avoid and I take all the rest" (Interview I.21)

In order to compare the quality of datasets, segments should be defined in the same way. This is possible for sociodemographic data, which rely on generic and objective descriptions (e.g. age, gender). This appears to be significantly more complex for data that are qualified in a subjective way, like preferences, since no standards for definitions are currently available. An exhaustive, precise and objective definition of every type of preferences appears to be unrealistic. In the same vein, as the data composing each segment are of heterogeneous natures, each segment should only be audited separately in order to reach a perfect assessment. Targeting methodology must then be checked independently for each attribute.

> "The advantage with sociodemographic is that there is no debate on definition […]. On behavioral data, it is more difficult […] we see that we cannot get everyone to agree" (Interview I.1)

Our results indicate that data are useful in different ways, depending on their nature and the information they convey. The dimensions of qualities that are sought by data buyers might differ in this way. For instance, as sociodemographic data concerns, for a part, long-term characteristics (e.g. age, gender), accuracy will be a striking dimension of quality. On the contrary, purchase intentions can rather be relevant in the short-term (e.g. need for ink-refill for printer), thus requiring freshness from the data.

In addition, measurements of data quality will thus target different dimensions that might be complementary but only partially relevant. First, due to heterogeneity in the dimensions of quality and their measurements, it may appear highly costly for market actors to produce global and exhaustive audits of datasets used in targeted advertising campaigns. Then, the fact that all dimensions of quality are not always relevant for data buyers can advocate for partial evaluations.

Monetizing and exchanging data

The monetization of data is operated through two main kinds of selling modes. First, data can be commoditized, sold as a standalone product by operators like data brokers (Crain 2018). According to the US Federal Trade Commission, *data broker*s are "companies whose primary business is collecting personal information about consumers from a variety of sources and aggregating, analyzing, and sharing that information, or information derived from it " (FTC 2014, p. 3). Gu, Madio, and Reggiani (2019) distinguish three types of data brokers on the basis of this broad definition:

(6) business-to-business players whose core business is to collect and sell information on consumers without having any contact with them (e.g. Axciom, Datalogix)
(7) business-to-business players focusing on property and financial information (e.g. eBureau, Corelogic)

(8) platform players whose model is funded through the exploitation of consumer data, but based on the supply of another type of service to consumers (e.g. Facebook, Google, Amazon)

In this vision, data is an asset that represents business opportunities for companies that own customer databases, or websites that collect information about their visitors. Practically, data are most of the time sold under the form of audience segments, which means that web users are classified under categories and subcategories according to their sociodemographic characteristics, their preferences or purchase intentions. The library of categories and subcategories is called taxonomy.

> "We make our own mix to create interesting segments for advertisers […] We conceived originally a rather large index, with approximately 2,000 segments" (Interview I.6)

Data vendors are usually selling audience segments depending on the strengths of their assets. For instance, telecom operators could offer to value the data they collect in the context of their services, relying on fine and wide geolocation or business-to-business data. Another example is a French start-up we interviewed, specialized in the collection of data related to e-sport amateurs and thus providing refined audience segments in this category. Such segments usually contain profiles, often under the form of cookies, and are sold at a defined Cost per Thousand Impressions (CPM). Certain ad tech intermediaries called "onboarders" propose to operate a match between first and second or third party data, thus increasing the available information on identified profiles.

> "A web user, in average… we are around three to four cookies" (Interview I.2)

A type of tools emerged in since the mid-2010s to help advertisers in the classification of their own data (i.e. 1st party) and correlate them with exogenous data (i.e. 2nd party or

3rd party): they are named 'Data Management Platforms' (DMP) or 'Customer Data Platform' (CDP) when integrating CRM sources and use cases (eMarketer 2018). Advertisers can create and manage their own taxonomy, in order to tailor the audience they want to address through campaigns. When they purchase data from external sources, these data are enriching the existing database. DMPs functions are triple: integration (which also means matching profiles with their digital identities), analytics and activation (Elmeleegy et al. 2013). Activation means that DMPs are connected to intermediaries called 'Demand-Side Platforms' (DSP), softwares-as-a-service that are used by advertisers to manage and optimize their biddings in programmatic campaigns, but also enables to target the right digital profiles, corresponding to the expected audience (Yuan, Wang, et Zhao 2013). However, these tools are complex, costly to develop and implement and have features that often exceed firms' necessities or abilities. The development of these DMP signals a trend in the willingness to centralize data sources and produce consistent audience on this basis.

> "We do not make pay for a full DMP license, because a full one is costly, it takes a lot of time to implement and it requires teams that know how to use it" (Interview I.12)
> "4 or 5 years ago, everybody asked for DMP, DMP, DMP but those guys did not know what it was. They invested hundreds of millions of euros in DMPs, whereas they had only 10,000 lines in their customer base" (Interview I.7)

The other selling mode is to bundle data and media. Firms like Amazon, Facebook or Google are not selling direct access to their audience segments, but provide already segmented media inventories. These firms have proprietary tools (e.g. Custom Audience for Facebook) that enable to upload 1st party data from advertisers in order to match audiences with segment provided by such platforms. In doing so, platform make sure that the only access to their users segments can be done through media purchase. Such absence of reciprocity in the sharing of data is encapsulated in the notion of "walled garden", usually applied to these firms (Fulgoni 2018).

In this vision, data is an asset that helps to strengthen the competitiveness of the inventory, based on the fact that Google, Facebook or Amazon have massive audiences coming from various touchpoints (e.g. Google collects data from Gmail, Youtube, and many other services, mostly provided for free in B2C context). The choice of such selling method is thus explained by the capacity to reach very much larger audiences than their competitors. This result in the fact that Google and Facebook have constituted a duopoly over time in the online advertising market - notably in France where they concentrate 78% of the ad spending (Alix 2018) - and are progressively threatened in their position by Amazon. Data is therefore not directly 'assetized' here, but considered as a component in the provision of a global service. This brings simplicity in the buying process and prevents the use of data outside a selected inventory, which reinforces the value of this very data.

> "Ad spendings have been massively driven toward platforms like Facebook and Google, because they offer scale, precision and a one-stop shop […]. You cannot buy Facebook anywhere else than through Facebook" (Interview I.6).

> "You have to distinguish two main types of [data] providers: GAFA and the others because it is true that for the moment, the ones that succeed in obtaining the most voluminous and structured data on consumers and prospects intentions, globally speaking, are Google, Facebook and Amazon" (Interview I.3)

Beyond monetization, data can also be bartered through second-party agreements. Second-party data is the first-party data of another company that is provided in the context of partnerships on specific audiences. Alliances based on this trading are emerging in the French market, for instance in the clothes sector in order to enrich the collective knowledge about consumers' preferences (De Matharel 2019). These various types of monetization or assetization of data do not form separate markets in the sense that firms are not providing the same service even if relying on data. Yet, advertisers can mix data from various sources: tools like DMPs can help them to gather data from several sources, unify identities across devices and organize consistent segments.

> "Second-party data […] is a one-to-one agreement where we know exactly from where the data comes from, how it was collected, what it refers to" (Interview I.10)

Data selling modes may play important role on the ability to check database quality. When data are sold bundled to media, the access to the database is not necessarily offered, thus potentially preventing ex-ante assessments. This is particularly the case for *walled gardens* operators. Concerning ex-post, assessing quality ex-post may be more difficult as it may be prone to noise (i.e media characteristics) and only measurable towards advertising efficiency. Conversely, assessing the quality of a database independently from media may allow for more flexibility, as it can be tested according to different measures, ex-ante and ex-post. In this case, every type of assessment can focus on different dimension of quality.

It is also significant to underline that data vendors can refuse to let advertisers run ex-post tests on their campaign, as they prefer to provide proprietary performance measurements only. This was the case for Facebook until 2016 (Uchôa-Lefebvre 2016) and for Google until 2018 (Offremedia 2018). This recent change could be perceived as an answer to the advertisers that pleaded for more transparency, through the voice of the French subsidiary of the World Federation of Advertisers, l'Union des Annonceurs (La Tribune 2017; Jaimes 2017). In this sense, the provision of additional information on quality through assessments appears to be crucial for advertisers, as they will be potentially able to negotiate the price of the data or the media enriched with data, depending on quality. The ability to value data from both sides of the market thus depends on this capacity to provide third-party measurements that gives objective signals of quality.

Data quality for performance vs. Data quality for itself

As recalled by Schultz, Block, and Viswanathan (2018, p. 763), media planning and buying activities have always focused on one goal: deliver the right message to the right audience at the right time at the lowest possible cost. Online advertising is characterized since its beginning by a tension between to two paradigms. One – *branding* advertising – mimics offline media and focusing on the visibility of the campaign by a selected audience, usually priced at the CPM (Cost per Thousand views). Another one – *performance* advertising – uses analytics and technical capabilities of online media to determine the interactions (e.g. clicks, leads, conversion) of the web users with the ad content, and is usually priced in accordance (e.g. cost per action, cost per click, cost per lead) (Ouakrat, Beuscart, and Mellet 2010). The constant move toward finer measurements has been driven by firms' extended tracking capabilities that allows for more precise analytics. This translates the willingness to assess the effectiveness of campaigns and determine ‘which half of the money is wasted’.

For now, two approaches related to campaign quality exist on the online advertising market. The first one relies on *ex post* tests to measure the impact of the campaign after its conduct (de Baynast et Lendrevie 2014). The choice of metrics will depend on the objectives of the campaign (e.g. increase online sales, offline sales, customer commitment). This measurement is asked by advertisers, who want to increase their Return on Investment (ROI), and monitored by agencies. Measurements can be either site-centric, delivered by the publisher, or user-centric, based on panels and delivered by specific intermediaries. Among user-centric solutions in the French market, two main solutions exist: Nielsen DAR (Digital Ad Ratings) by Mediamétrie/NetRatings and Validated Campaign Essentials (vCE) by ComScore. These tools analyze the global performance (e.g. targeting, viewability, impressions) of a campaign based on the behavior of a panel followed across the web. The score provided – on-target coverage – thus measures to what extent the campaign had been viewed by the targeted audience through a comparison with a sociodemographic panel (Mercanti-Guérin and Vincent 2016). In that sense, these advertising effectiveness measurements are also ex-post solutions for data quality assessment, as they help to determine if the data used for targeting was efficient in terms of performance.

> "We validate the relevance of the data we use by performing tests with an indicator called DAR, which eventually helps us to check if the promise of the third-party data we bought is respected […] We know [then if] we will work with this provider or another one" (Interview I.4)

Such solutions are holistic, as they do not provide assessment on the detailed qualities of used data, but a global vision of their usefulness in the objective of the campaign. However, *ex-post* solutions are convenient for advertisers who favor a ROI approach, over an intrinsic interest for data quality. In this context, it appears that the quality expected and evaluated by *ex post* solutions is 'fitness for use', as the process is directed toward the evaluation of a global performance where data is used as targeting mean.

Considering *performance* campaigns, it appears that efficiency assessment is directly measured by transformation from displaying ads to purchase, which rules out the use of such ex-post measurement solution. However, marketers still use the measurement solution in pre-test campaigns to understand what services to use for the real one. In this case, *ex-post* solutions provide them with metrics to position their campaign. Turning now to *branding* campaigns, it appears that *ex-post* measurement is extremely useful to understand if intended targeting has been effectively performed, which therefore allow to approximate efficiency.

Conversely, the second approach of campaign quality takes place *ex-ante,* where advertisers verify campaign conditions before it runs, trying to minimize expected risk associated with running ads and targeting users. A first *ex-ante* solution for data quality assessments on the French market is provided by Médiamétrie, one of the leading firms in audience measurement. It leverages a representative panel of the French population to provide scores for datasets used in targeted advertising. However, such solution is only able to assess segments of simple characteristics (i.e. sociodemographic). More complex datasets - providing segments related to purchase intent for example - may not be measurable by such methodology, due to a lack of information from the panel.

A second *ex-ante* solution is provided by the Centre d'Etudes des Supports Publicitaires (CESP), a French trade association auditing audience and media research solutions. It intends to provide a comprehensive analysis of the data collection process as well as the segments construction methodology. The CESP thus produces recommendations associated to a theoretical appreciation of the dataset quality. However, this solution does not allow for empirical assessment of quality. Both solutions appear to be desirable for *branding* and *performance* campaigns as it allows advertisers to better direct their investments *prior* to the start of their campaigns. More generally, ex-ante solutions leverage campaign characteristics to assess a quality score used to predict theoretical efficiency. Score is produced in auditing database characteristics and methods used by firms to produce targeting. Such audit, relates to an effort from data vendors to ensure that data and its use for targeting has characteristics that minimizes risk of poor performance. In that sense, data vendors can signal their good quality through ex-ante assessments.

> "There is quite strong demand from the market on quality of sources, which is to see from where comes and what is the recency and frequency of these data" (Interview I.5)

> "By offering this transparency, we are identified as a trusted third party and serious people, the fact of having being audited […] also helps to gain credibility" (Interview I.6)

Running ex-ante assessments can also be beneficial for advertisers that collect first party data. Advertisers that do so can get information on the quality of their own database. This interest can be described as a form of "data hygiene". For these data-savvy advertisers, managing properly a first-party database is important to enhance their targeted campaigns. However, this situation remains relatively scarce, as significant human resources and investments are needed to internalize such data management capacities. This is worthy for advertisers that are in capacity to tailor specific audiences and know properly the segments they want to address. It can also be important for the valuation of their data, especially if their trade or exchange them, like in the case of second party agreements. Additionally, it is also a way to ensure that their data collection is well done and error-proof, thus reinforcing trust between the brand and their customers, notably in terms of respect for privacy. With ex-ante assessments, data vendors can thus receive certifications of quality.

> "On our database, which is permanently audited, we have teams that are dedicated to reconcile all our databases" (Interview I.2)

Ex-ante and *ex-post* measurement technologies thus produce different definitions of quality. However, such definitions are not opposed to one another and may even appear as complementary. Practically, we see that it is not the case on the market and ex-post technique is more widespread than ex-ante ones.

A first reason lies in the fact that firms of the online advertising duopoly, Google and Facebook, only accept ex-post techniques. A second reason may be found in the fact that advertisers may not always be able to determine ex-ante on which segment their campaign is more likely to perform. Targeting decisions are not necessarily carefully forecasted, but adjusted almost in real time, depending on daily variations. In this case, performing tests using ex-post techniques allows identifying the best segments for a solid targeting strategy. A third reason may also lie in the fact that ex-post techniques are more mature, already implemented on the market and seems to be perceived as a substitute to ex-ante techniques, which is not exactly the case. We sum-up the difference between techniques and their market impact in the table 4 below. While ex-post techniques are widespread on the market, ex-ante ones - while just introduced - seem to exhibit a lower adoption rate. In this case, the market may privilege "test and adjust" solutions (ex-post) in the long run.

Solution	Principles	Type of Campaigns	Pros	Cons	Theoretical Impact	Real impact	Dimensions of quality assessed
Ex-ante	Use characteristics to build a theoretical efficiency	Branding and performance	Signal a theoretical quality before using the service	Not linked to empirical evidence + non exhaustive	Help advertisers choose their services **before** using them	Limited willingness to pay due to substitution with Ex-post services	**Médiamétrie**: accuracy **CESP**: GDPR-compliance, freshness, currency, integrity.
Ex-Post	Monitor ongoing campaign an adjust it after results		Monitor real quality	Part of the campaign is wasted to establish true quality	Limit mistakes in repeated interactions	Full Impact. Ex-post is used to monitor performance	Fitness for use

Table 4. Characteristics of ex-ante and ex-posts data quality assessment techniques.

Data quality assessment and market competition

In a context where data is a commodity, monetized for itself or providing additional value to media for advertising, the question of quality has an influence of the competitiveness of market actors. For a data vendor, signaling the good quality of data can appear as a mean of differentiation, and thus help to gain competitive advantage (Ross et Shetty 1985).

This differentiation strategy can be used by data vendors of the competitive fringe towards the Google/Facebook duopoly. Indeed, these two firms have a massive access to data and are thus in position to provide the largest *reach* - the percentage of the population that will be exposed to the campaign - and give access to almost every types of audiences to advertisers. This high *reach* rate is helpful for advertisers, as there is a threshold to start running a campaign due to the fixed cost that are engaged. It is also useful for advertisers looking for niche audiences (e.g. specific products, categories) to be able to target sufficient amount of individuals.

> "Google, Facebook, Amazon have a strike force which is enormous. Their audience volumes are enormous, so we cannot miss this out" (Interview I.14)

Yet, in theory, the quantity of available data does not necessarily goes along with a good quality of data. As the competitive fringe of data vendors (e.g. data brokers, publishers) are not able to compete with the duopoly on the quantity of data, some of them are engaging in a race for quality and join their forces in this context. For instance, the "Alliance Gravity" is a group of firms composed of more than 150 French media groups, telco operators and retail brands selling together an access to 2,000 segments composed of an aggregation of data. The main idea is to get closer from Google and Facebook in terms of *reach* capacity and bet on quality to differentiate. They also want to be competitive on the question of transparency, assuming that advertisers are increasingly worried about the opacity surrounding data provision and billing activities. This alliance thus intends to provide a higher level of information for advertisers than the duopoly and other data brokers, especially regarding the various dimensions of quality, such as recency-frequency or GDPR-compliance, based on the results of the CESP audit. Trust in quality can thus become a differentiation factor for data providers and ex-ante tests are a way to signal it.

> "There has never been more distrust and suspicion regarding data and how it has been collected, shared, used" (Interview I.1)

> "When it is opaque, complicated, then we have to trust them [i.e. GAFA] so they have a greater ability to make hidden margins" (Interview I.7)

Another option for data vendors is to specialize in a specific type of audience or product to offer more precision and depth: this is the case for Join, for example, that focuses on e-sport audiences.

One main difficulty here is found on the cost of assessment. Considering ex-post solutions, the service provided by Nielsen or ComScore is paid, directly or indirectly, by the advertisers who are willing to have indications on the performance of the campaign. Such indication could help the agency or the advertiser - when the brand has internalized the media and data buying process - to negotiate the price of the campaign *a posteriori* with the vendor. Paying for ex-post assessments thus triggers potential discounts on the final price for media and data, depending on the performance of the campaign. On the contrary, ex-ante assessments are delivering certifications of good quality: in this context and from our interviews, it appears that advertisers are not willing to pay for the cost of this market intermediation, which is charged to the data vendor. Data vendors of the competitive fringe - Google and Facebook are not paying for *ex-ante* assessments - will thus see their margin hampered by such cost. All things considered, the incentive for signaling good data quality is partly compensated by this additional cost.

Eventually, a chicken-and-egg problem remains: is the offer for solutions is driving a new demand, or is a preliminary demand that fostered the emergence of solutions? This question is hard to answer in a definitive way. The providers of these assessment methods are playing a central role in this situation. These market intermediaries are, in a way, responding to a demand from the market actors. This evaluation activity is rooted in the history of advertising, as a mean of professionalization and legitimacy to the sector (Gaertner 2008).

Ex-post solutions providers, Nielsen and ComScore, are competing with a similar service for the same market. On the ex-ante facet, Médiamétrie and the CESP provide two different types of solutions focusing on different but complementary dimensions of quality. These four actors are interrelated: Nielsen's Digital Ad Ratings was launched in collaboration with Médiamétrie in 2013 and the CESP has audited comScore and Nielsen's solutions (Offremedia 2017), but also the ex-ante solution of Médiamétrie (Offremedia 2019). Beyond the relations among these organizations, it is also significant to underline that the boards of CESP (an association) and Médiamétrie (a private company) are constituted by firms of the advertising sector (i.e. trade associations, agencies, publishers, ad tech intermediaries). This intricacy among intermediaries reinforces the legitimacy of such organizations as evaluators for market practices.

> "The advertisers are pushing for [our ex-ante solution development] the most. I think it is because we are close from the publishers" (Interview I.1)

> "Our board agreed to let us investigate on what could be the needs of the advertisers [regarding ex-ante solutions]" (Interview I.5)

The coherence between the spread of these solutions and the discourse of trade associations on the need for more transparency and indicators tends to show that these intermediaries have been incentivized to expand their solutions to address these new issues. It is also, for Mediamétrie as well as for the CESP, a way to make a new use of their own assets, the sociodemographic panel for the former and the auditing methodology for the later.

Considering all these elements, it appears that these different definitions of quality can become a differentiation element providing a competitive advantage for market actors willing to signal their good quality. The provision of assessment solutions, especially ex-ante, is thus desirable for market actors in this context but is limited by its cost and by the fact performance indicators, for the moment, and *reach* capacities are driving the market, more than data quality *per se*.

Discussion

The online advertising market has been in a state of exponential growth since its invention, and recently started to invest in transparency. Such position has been endorsed by both small and big actors, with the desire to innovate towards more accountability regarding advertising spending. This is consistent with the historical trend depicted by Beniger (1986) on the control necessity for actors, especially at the moment of emergence of new technologies. The decentralized production of data and its spread all across the world are generating negative externalities that market actors are urged to limit in order to preserve the legitimacy and trust on the online advertising activity. In that sense, the massive spread of ex-post and the emergence of ex-ante solutions for data quality assessment are contributing to produce evidence of credibility in a particular context.

However, the competitive landscape is dominated by a few firms that provide massive reach in integrated solutions, which is a competitive advantage sufficient enough to be in a position of under-investment in accountability. This has been reproached to both Facebook and Google, notably by the French competition authority (Pépin 2018). To counter such domination, market actors from the competitive fringe have created alliances based on the gathering of inventories, data or tracking capabilities. They offer data as a distinct product and use transparency on quality as a competitive advantage. Insofar, data quality becomes a differentiation element and requires the use of tests to posit differences in qualities among datasets. Assessment solutions thus increase the number of assessable facets and contribute to the ability for data to be valued.

Assessment solutions, especially ex-ante as they shed light on various dimensions of quality, provide means of valuation of data that was not taken into consideration previously. In this sense, they can be considered as calculative agencies, which contribute to "make goods calculable" (Muniesa, Millo, and Callon 2007). The fact that data are more precisely valued through these assessments is rooted in the "assetization" movement concerning data (Beauvisage et Mellet 2020). Ex-post tests were evaluating data's "fitness for use" *de facto* through the evaluation of campaigns. Ex-ante tests are, on one side, isolating and valuing data *per se*, especially through the granting of certifications for data vendors. It reinforces the idea that data could be sold without a media bundle. On the other side, with the trend of *data hygiene,* these types of tests are assimilating data as valuable assets possessed by advertisers - assets they have to manage. Originally, such process was not obvious for certain advertisers, who felt being in the position of Moliere's Monsieur Jourdain, creating assets on a daily basis without even knowing it. Ex-ante quality tests are, in that sense, the most recent part of the evangelization process driven by market intermediaries.

In our case, market intermediaries are of the *evaluator* type as described by Bessy et Chauvin (2013) whose core activity consists in "producing evaluations, rankings or ratings" (p.101). Exploring new facets of the advertising product can be considered as a differentiation strategy that reduces competitive pressure. The potential success of the ex-ante tests is also linked to the reputation of these market intermediaries. They contributed, in the past decades, to build a more "professional" advertising sector (Gaertner 2008). The promotion of assessments for the quality of data thus contributes to anchor the assetization process in the evolution timeline of the market. Moreover, as their boards are composed of market actors, they are granted with an uncontested credit. Advertisers, publishers and data vendors will be more willing to trust these third-parties as they know, sometimes personally, these middlemen. In a context where programmatic advertising has automatized and partly dehumanized the online advertising value chain, having the possibility to trust incarnated intermediaries is perceived as more trustworthy.

These intermediaries are contributing to "clear the market" by providing infrastructures necessary for its functioning (Spulber 1996), such as certifications and evaluations. This contributes to make the market more reliable, but also to bring new opportunities for competition among actors. In this vision, ex-ante tests appear to be theoretically more cost-saving than ex-post tests in terms of transaction costs (Williamson 1985). Repeating tests for each campaign is less efficient, than granting once for all the collection method or the quality of a database that might be used several times.

However, as underlined by Bessy et Chauvin (2013), “the power of valuation of an intermediary can be measured through the effects of the valuations it produces, whatever their form or their logics may be” (p.103). As demonstrated in our paper, ex-ante assessments will have difficulties to overcome the performance paradigm that focuses only on a global outcome. There is a collective action problem here as the industry is not ready, collectively, to invest on these measurements. The fact that each test is addressing different dimensions of quality and the lack of standard for consumer preferences make even more difficult the adoption of a global solution. In this context, private voluntary solutions do not appear as the most efficient for resolving the quality issue on the online advertising market. Public regulations, like the GDPR on the respect of consumer consent, are thus more consistent and likely to succeed in reducing market failures.

Conclusion, recommendations and future research

Through this article, we analyze how the introduction of data quality assessment solutions reveals the different definitions of quality and could be desirable for market actors. We show that the heterogeneity in data natures and uses, as well as in data quality assessment solutions on the market impact both aspects. Indeed, the existing solutions are assessing various dimensions of quality, but in complementary way. Depending on the competitiveness of the data they monetize, data vendors can be incentivized to signal their quality. Yet, it appears from the interviews we have conducted that the desire for more information is not always present on the buying side.

Several recommendations can be made from our findings and discussion. First, it appears that the market intermediaries are relevant to operate clarifications on the definitions of quality. Yet, ex-ante tests still have to demonstrate their usefulness for data buyers, practically speaking. This is even more necessary as the existing solutions are not covering the whole types of data. In the long run, the standardization of categories (e.g. user preferences, taxonomies) could help to extend the scope and the efficiency of ex-ante solutions. It could give further possibilities for data buyers to compare the characteristics of the offers. Another breakthrough would be the unification of ex-ante solutions in order to provide more global services.

On the side of data vendors, the alliance strategy of the competitive fringe operators is relevant as it helps to acquire a reach rate almost comparable to Google and Facebook. In addition, the specialization of certain actors (i.e. focusing and selling data on specific vertical categories) appears to be interesting as it could lead to a kind of niche monopolization, thus gaining comparative advantages on global actors.

These grievances need to be replaced in their context. The automation of the allocation process in the industry brought a technical complexity, in comparison with human-based interactions of the offline era. Beyond, the growing concern of web users about the use of their personal data (i.e. data that can link the information to the person's identity), often described under the term privacy (Acquisti, Brandimarte, et Loewenstein 2015), is fueling debates about targeted advertising and gave paths to jurisprudences and new regulations, such as the GDPR (Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data) in the European Union.

This article contributes to a stream of literature dedicated to the understanding of modern advertising market mechanisms and technologies. Yet, our work is nurtured and limited by its methodology, as interviews are not flawless to picture the global state of an industry. In this sense, quantitative works could, for instance, explore the concrete impact of data quality on the advertising efficiency. As well, a modelling process could help to encapsulate the different dimensions of quality and contribute to provide anew, and more consistent solution for data quality assessment. On the side of competition and antitrust issues, this work could pave the way for further research on the way data, as assets, are concretely contributing to reinforce the market power of the online advertising duopoly.

References

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. 2015. « Privacy and human behavior in the age of information ». *Science* 347 (6221): 509-514.

Akerlof, George A. 1970. « The Market for" Lemons": Quality Uncertainty and the Market Mechanism ». *The Quarterly Journal of Economics* 84 (3): 488-500.

Alix, Christophe. 2018. « La pub hexagonale, un duopole Google-Facebook ». *liberation.fr*, janvier 26. https://www.liberation.fr/futurs/2018/01/26/la-pub-hexagonale-un-duopole-google-facebook_1625265.

Beauvisage, Thomas, and Kevin Mellet. 2020. « Datassets: Assetizing and Marketizing Personal Data ». In *Assetization: Turning Things into Assets in Technoscientific Capitalism.*, MIT Press.

Beckert, Jens, and Christine Musselin. 2013. *Constructing quality: The classification of goods in markets*. OUP Oxford.

Beniger, James R. 1986. *The Control Revolution: Technological and Economic Origins of the Information Society*. Harvard University Press.

Bessy, Christian, and Pierre-Marie Chauvin. 2013. « The power of market intermediaries: From information to valuation processes ». *Valuation studies* 1 (1): 83-117.

Bourdon, Jérôme, and Cécile Méadel. 2014. *Television audiences across the world: Deconstructing the ratings machine*. Springer.

Brousseau, Eric, and Jean-Michel Glachant. 2014. *The manufacturing of markets: legal, political and economic dynamics*. Cambridge University Press.

Cai, Li, and Yangyong Zhu. 2015. « The Challenges of Data Quality and Data Quality Assessment in the Big Data Era ». *Data Science Journal* 14 (0): 2.

Callon, Michel, Cécile Méadel, and Vololona Rabeharisoa. 2002. « The economy of qualities ». *Economy and society* 31 (2): 194-217.

Chamberlin, Edward H. 1953. « The product as an economic variable ». *The Quarterly Journal of Economics* 67 (1): 1-29.

Chamberlin, Edward H. 1933. « The Theory of Monopolistic Competition ». *Harvard Economic*.

Corbin, Juliet M, and Anselm Strauss. 1990. « Grounded theory research: Procedures, canons, and evaluative criteria ». *Qualitative sociology* 13 (1): 3-21.

Crain, Matthew. 2018. « The limits of transparency: Data brokers and commodification ». *New Media & Society* 20 (1): 88-104.

Creswell, John W. 2007. *Qualitative Inquiry & Research Design*. Sage Publications, Inc.

Crosby, Philip B. 1979. *Quality is free: The art of making quality certain*. New York: McGraw-Hill.

de Baynast, Arnaud, and Jacques Lendrevie. 2014. *Publicitor. Publicité online & offline.* 8ème édition. Dunod.

De Matharel, Lélia. 2019. « La galaxie Mulliez crée Fashion Data, alliance data anti-Amazon d'enseignes de mode ». *lsa-conso.fr*. https://www.lsa-conso.fr/pingki-houang-fashion3-nous-creons-fashion-data-alliance-data-anti-amazon-d-enseignes-de-mode,306671.

Del Frate, Yves. 2020. « Votre data est-elle bio? » *e-marketing.fr*. https://www.e-marketing.fr/Thematique/data-1091/Breves/Votre-data-est-elle-bio-345603.htm#.

Demaziere, Didier, and Claude Dubar. 1997. « EC Hughes, initiateur et précurseur critique de la Grounded Theory ». *Sociétés contemporaines* 27 (1): 49-55.

Digiday Editors. 2018. « The quality of vendor data is shocking': Overheard at Digiday Publishing Summit Europe ». *Digiday.com*, octobre 17. https://digiday.com/media/quality-vendor-data-shocking-oh-digiday-publishing-summit-europe/.

Downie, Jason. 2018. « Facebook Proves It: Data Quality, Not Scale, Matters the Most ». *Adweek.com*, mai 8. https://www.adweek.com/digital/facebook-proves-it-data-quality-not-scale-matters-the-most/.

Ebeling, Mary FE. 2016. *Healthcare and big data*. Palgrave Macmillan.

Elmeleegy, Hazem, Yinan Li, Yan Qi, Peter Wilmot, Mingxi Wu, Santanu Kolay, Ali Dasdan, and Songting Chen. 2013. « Overview of turn data management platform for digital advertising ». *Proceedings of the VLDB Endowment* 6 (11): 1138-1149.

eMarketer. 2016. « Data Quality Becomes a Top Concern for Marketers. Poor-quality intelligence leads to lost sales, inefficiency. » *emarketer.com*. https://www.emarketer.com/Article/Data-Quality-Becomes-Top-Concern-Marketers/1014751.

eMarketer. 2018. « Understanding the "Data" in Customer Data Platforms ».

eMarketer. 2019. « Programmatic Digital Display Ad Spending in France, 2016-2020 ».

Farahat, Ayman, et Michael C Bailey. 2012. « How effective is targeted advertising? » In *Proceedings of the 21st international conference on World Wide Web*, 111-120. ACM.

Fernández-Barcala, Marta, Manuel González-Díaz, and Emmanuel Raynaud. 2014. *How to manufacture quality: the diversity of institutional solutions and how they interact in agrifood markets*.

Forrester. 2019. *Why Marketers Can't Ignore Data Quality.*

FTC. 2014. *Data Brokers. A call for Transparency and Accountability*.

Fulgoni, Gian M. 2018. « How Limited Data Access Constrains Marketing-Mix Analytical Efforts: Why Data Barriers Are Preventing Marketers From Optimizing Marketing Spend ». *Journal of Advertising Research* 58 (4): 390-393.

Gaertner, Laure. 2008. « Que produisent les publicitaires? Retour socio-historique sur la formation d'une expertise ». *Management & Avenir*, nº 1: 140-155.

Gao, Jerry, Chunli Xie, and Chuanqi Tao. 2016. « Big Data Validation and Quality Assurance--Issuses, Challenges, and Needs ». In , 433-441. IEEE.

Garvin, David A. 1984. « What Does "Product Quality" Really Mean ». *Sloan management review* 25.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine. Chicago.

Goldfarb, Avi, and Catherine Tucker. 2008. « Search engine advertising: Pricing ads to context ».

Goldfarb, Avi, and Catherine Tucker. 2011. « Chapter 6 - Online advertising ». In *Advances in Computers*. Vol. 81. Elsevier.

Goldfarb, Avi, and Catherine Tucker. 2019. « Digital Economics ». *Journal of Economic Literature* 57 (1): 3-43.

Goulding, Christina. 1998. « Grounded theory: the missing methodology on the interpretivist agenda ». *Qualitative Market Research: an international journal* 1 (1): 50-57.

Gu, Yiquan, Leonardo Madio, and Carlo Reggiani. 2019. « Data brokers co-opetition: Why do data brokers share data in some markets and compete in others? »

Jaimes, Nicolas. 2017. « Facebook doit se faire auditer par le CESP sans attendre ». *JournalDuNet.com*, juillet 10.

https://www.journaldunet.com/ebusiness/publicite/1196311-jean-luc-chetrit-union-des-annonceurs/.

Jaye, Daniel. 2019. « Consent Fraud: A Simmering Problem That Could Scald The Ecosystem ». *Adexchanger.com*, février 26. https://www.adexchanger.com/data-driven-thinking/consent-fraud-a-simmering-problem-that-could-scald-the-ecosystem/.

Jones, Philip, and John Hudson. 1996. « Standardization and the costs of assessing quality ». *The economics of standardization* 12 (2): 355-361. doi:10.1016/0176-2680(95)00021-6.

Joseph, Seb. 2019. « Confessions of a location data exec: 'It's a Ponzi scheme' ». *Digiday.com*, février 26. https://digiday.com/marketing/confessions-location-data-exec/.

Karpik, Lucien. 1989. « L'économie de la qualité ». *Revue française de sociologie*, 187-210.

Kumar, V., and Shaphali Gupta. 2016. « Conceptualizing the Evolution and Future of Advertising ». *Journal of Advertising* 45 (3): 302-317.

Lakshen, Guma Abdulkhader, Sanja Vraneš, and Valentina Janev. 2016. « Big data and quality: A literature review ». In *2016 24th telecommunications forum (TELFOR)*, 1-4. IEEE.

Lambrecht, Anja, and Catherine Tucker. 2013. « When does retargeting work? Information specificity in online advertising ». *Journal of Marketing Research* 50 (5): 561-576.

Lancaster, Kelvin J. 1966. « A new approach to consumer theory ». *Journal of political economy* 74 (2): 132-157.

LaTribune. 2017. « Publicité numérique : les annonceurs demandent plus de transparence ». *LaTribune.fr*, décembre 8. https://www.latribune.fr/technos-medias/internet/publicite-numerique-les-annonceurs-demandent-plus-de-transparence-760999.html.

Leland, Hayne E. 1979. « Quacks, lemons, and licensing: A theory of minimum quality standards ». *Journal of political economy* 87 (6): 1328-1346.

Loshin, David. 2010. *The practitioner's guide to data quality improvement*. Elsevier.

Mercanti-Guérin, Maria, and Michèle Vincent. 2016. *Publicité digitale*. Dunod.

Merino, Jorge, Ismael Caballero, Bibiano Rivas, Manuel Serrano, and Mario Piattini. 2016. « A data quality in use model for big data ». *Future Generation Computer Systems* 63: 123-130.

Miles, Matthew B, and A Michael Huberman. 2003. *Analyse des données qualitatives*. De Boeck Supérieur.

Muniesa, Fabian, Yuval Millo, and Michel Callon. 2007. « An introduction to market devices ». *The sociological review* 55 (2_suppl): 1-12.

Nelson, Phillip. 1970. « Information and consumer behavior ». *Journal of political economy* 78 (2): 311-329.

Nelson, Phillip. 1974. « Advertising as information ». *Journal of political economy* 82 (4): 729-754.

Offremedia. 2017. « Les conclusions de l'audit du CESP sur les solutions DAR (Nielsen) et vCE (comScore) ». *offremedia.com*. https://www.offremedia.com/les-conclusions-de-laudit-du-cesp-sur-les-solutions-dar-nielsen-et-vce-comscore.

Offremedia. 2018. « Les 23 partenaires de mesure 3rd-party du programme «Google Measurement Partners» ». *offremedia.com*. https://www.offremedia.com/les-23-partenaires-de-mesure-3rd-party-du-programme-google-measurement-partners.

Offremedia. 2019. « Les conclusions de l'audit du CESP sur la solution Data Checking de Médiamétrie ». *offremedia.com*. https://www.offremedia.com/les-conclusions-de-laudit-du-cesp-sur-la-solution-data-checking-de-mediametrie.

Olenski, Steve. 2018. « 3 barriers to Data Quality and How to Solve For Them ». *Forbes.com*, avril 23. https://www.forbes.com/sites/steveolenski/2018/04/23/3-barriers-to-data-quality-and-how-to-solve-for-them/.

OnAudience.com. 2018. *Global Data Market Size 2017-2019*.

Ouakrat, Alan, Jean-Samuel Beuscart, and Kevin Mellet. 2010. « Les régies publicitaires de la presse en ligne ». *Réseaux* 160-161 (2-3): 133-161.

Pépin, Guénaël. 2018. « Face à Facebook et Google, l'Autorité de la concurrence veut rééquilibrer le marché publicitaire ». Nextinpact.com. https://www.nextinpact.com/news/106260-face-a-facebook-et-google-autorite-concurrence-veut-reequilibrer-marche-publicitaire.htm.

Rao, Akshay R, and Kent B Monroe. 1989. « The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review ». *Journal of marketing Research* 26 (3): 351-357.

Reeves, Carol A, and David A Bednar. 1994. « Defining quality: alternatives and implications ». *Academy of management Review* 19 (3): 419-445.

Ross, Joel E, and Y Krishna Shetty. 1985. « Making quality a fundamental part of strategy ». *Long Range Planning* 18 (1): 53-58.

Schultz, Don E, Martin P Block, and Vijay Viswanathan. 2018. « Consumer-driven media planning and buying ». *Journal of Marketing Communications* 24 (8): 761-778.

Sebastian-Coleman, Laura. 2012. *Measuring data quality for ongoing improvement: a data quality assessment framework*. Newnes.

Shapiro, Carl. 1986. « Investment, moral hazard, and occupational licensing ». *The Review of Economic Studies* 53 (5): 843-862.

Spence, Michael. 1978. « Job market signaling ». In *Uncertainty in Economics*, 281-306. Elsevier.

Spulber, Daniel F. 1996. « Market microstructure and intermediation ». *Journal of Economic perspectives* 10 (3): 135-152.

Stigler, George J. 1961. « The economics of information ». *Journal of political economy* 69 (3): 213-225.

Strauss, Anselm, and Juliet Corbin. 1994. « Grounded theory methodology ». *Handbook of qualitative research* 17: 273-285.

Suddaby, Roy. 2006. « From the Editors: What Grounded Theory Is Not ». *The Academy of Management Journal*, 633-642.

Turow, Joseph. 2012. *The daily you: How the new advertising industry is defining your identity and your worth*. Yale University Press.

Uchôa-Lefebvre, Luciana. 2016. « Visibilité : comScore, Nielsen et IAS signent avec Facebook ». *ad-exchange.fr*. http://ad-exchange.fr/visibilite-comscore-signe-un-partenariat-mondial-avec-facebook-30221/.

Velkar, Aashish. 2012. *Markets and measurements in nineteenth-century Britain*. Cambridge University Press.

Velkar, Aashish. 2014. « 2. Measurement systems as market foundations: perspectives from historical markets ». In *The Manufacturing of Markets: Legal, political and economic dynamics*, 17-36. Cambridge University Press.

Wang, Richard Y, and Diane M Strong. 1996. « Beyond accuracy: What data quality means to data consumers ». *Journal of management information systems* 12 (4): 5-33.

Williamson, Oliver E. 1985. *The Economic Intstitutions of Capitalism*. Simon and Schuster.

Yuan, Shuai, Jun Wang, and Xiaoxue Zhao. 2013. « Real-time bidding for online advertising: measurement and analysis ». In , 3. ACM.

Zins, Chaim. 2007. « Conceptual approaches for defining data, information, and knowledge ». *Journal of the American society for information science and technology* 58 (4): 479-493.

Chaire Gouvernance et Régulation
Fondation Paris-Dauphine
Place du Maréchal de Lattre de Tassigny - 75016 Paris (France)
http://chairgovreg.fondation-dauphine.fr