



Using Big Data for Online Advertising Without Wastage: Wishful Dream, Nightmare or Reality?

Mark Grether

KEYWORDS

*Online Advertising, Big Data,
Third-party Data Provider,
Personalization,
Wastage, Retargeting*

•

THE AUTHOR

Mark Grether,
Global Chief Operating Officer,
Xaxis, New York City , US,
mark.grether@gmail.com

Big Data and Online Advertising: High Expectations ///

Everyone right now is talking about big data, which is viewed as the next great innovation challenge for marketers. The digitalization of the entire advertising industry is generating ever increasing amounts of data that must be collected, analyzed and interpreted. Using real-time and comprehensive data-assisted decision making, companies are hoping for significant competitive advantages by improving processes and creating more options for tailoring and personalizing services. Digital advertising is an important application for this personalization. Customized advertising will be more effective, cost less, and be better received by society. Companies like Google and Facebook are playing a vanguard role here, while their stock values demonstrate the economic potential that can be realized through big data.

A New Data Market Emerges ///

But advertisers are not the only ones who have high expectations for the possibilities of big data. Progress in database and analytics systems has opened the doors to this new business opportunity to more and more smaller companies. Figure 1 shows some of the players who are active in this new big data market. Particularly in online advertising, many companies are trying to develop their own business ideas and claim a piece of the growing online advertising pie by using big data tools. In digital jargon, these companies are called third-party data providers. They are transferring the profitable data business, which large market research institutes like GfK, TNS, Nielsen and Comscore have established in the non-digital environ-

FIGURE 1:
The Big Data Vendor Landscape



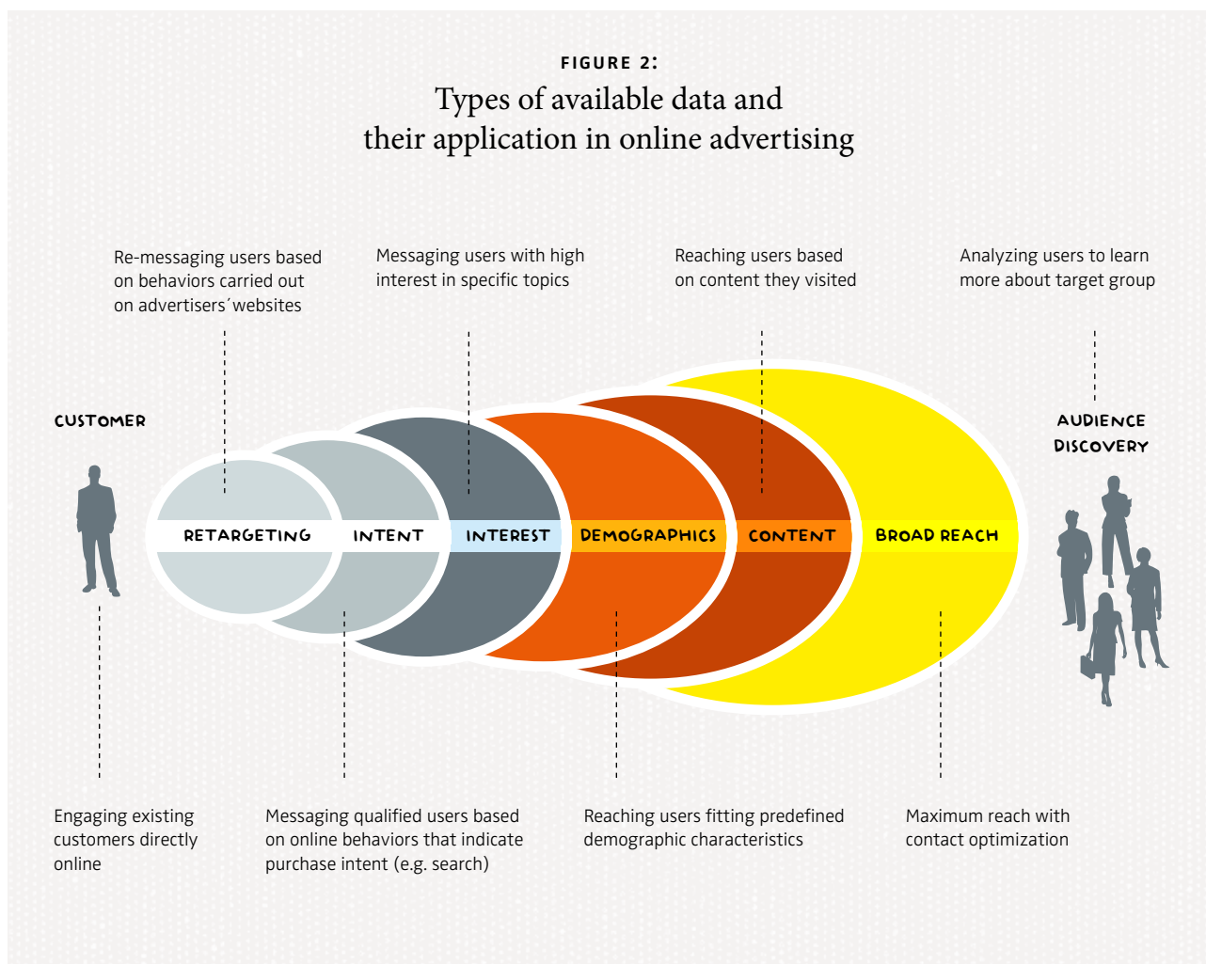
Source: <http://www.mammothdb.com/the-big-data-landscape-by-mammothdb/>

ment, to the digital world – user data that has been acquired through the analysis of user behavior on the internet, is sold and licensed. But the undeniable opportunities are offset by substantial risks. The profitable use of big data is not without pitfalls, and some of the business models on the fringes of big data and online advertising simply do not work. These fledgling companies rarely succeed in achieving a competitive advantage in the market. Many are fighting for economic survival.

Using Big Data to Optimize Advertising /// The big promise of big data to advertising is improved accuracy of communication. Advertising is expected to become more relevant and less expensive as a result of less wastage. Different data is necessary depending on individual advertising goals. Basically, advertising activities can either be performance-related or support the brand image. The greater the focus on immediate sales success, the more data is needed to promote individual customer contact and re-targeting. But if increasing brand recognition is what matters, the focus will be more on general interest data and nonspecific messages (see Figure 2).

> **Big Data and Performance Marketing** /// With performance marketing, advertising is billed solely on the basis of the performance of an agreed action by a targeted online user. In simplest terms, this action is a click on the advertisement. Now, with big data, it is possible to investigate which variables influence this click. Predictive re-targeting methods are frequently used to this end. These are based on data on users who were already very close to taking the desired action. Data mining tools are used to search for other users who have matching behavioral profiles. Searching for these statistical twins involves enormous amounts of data with diverse quality in different sets, often exhibiting large gaps. Furthermore, the data must be evaluated within milliseconds. But if a sufficient number of such twins can be found, reach and click rates can be increased significantly. For example, if a campaign with a cost-per-thousand price of €1 achieves a factor of five, the value of the underlying algorithm is €5 (per thousand). The value of the advertising space therefore increases five-fold.

FIGURE 2:
Types of available data and
their application in online advertising



> **Big Data and Branding** /// Branding campaigns frequently aim to improve brand image or recognition. This is traditionally a domain of TV advertising. Therefore, the online advertising world has adopted indicators like net reach or gross rating points from TV advertising. The success of a branding campaign is judged by maximum contact with a given target group. In many cases, sociodemographics like age and gender determine the relevant segments. Data mining is used to make a valid prediction of these characteristics for as many online users as possible. Usually, the greater the reach, the less precise the forecasting of characteristics, and this is a trade-off that must be considered. Provided that the data is valid, an advertiser can significantly reduce its media costs this way. Advertising is delivered only to its target group, driving down wastage significantly. A good example of this type of data usage is Facebook.

With the login data of its users, Facebook has access to well-validated age and gender information, and it achieves enormous reach via various devices. The underlying data is ideal for the precise delivery of advertising to the target group. Particularly with video advertising, which is primarily used to increase brand recognition, age and gender are well-suited criteria for targeting.

The Pitfalls of Monetizing Big Data in Advertising ///

What looks deceptively simple through its success is frequently quite difficult to implement in practice. This holds particularly true for the aforementioned third-party data providers, who unlike Facebook or Google do not use their own data, but live on the sale of such data. Aside from data quality, the biggest problem lies in determining a reasonable price for the data. The box on the side describes why it is difficult to determine the value and the quality of data.

{ Box 1 }

THE PROBLEM WITH VALUING DATA FOR ONLINE ADVERTISING

The Value of Data Varies with Different Applications

Specific information about online users, for instance whether they are male or female, interested in finance or sports, or live in New York or Los Angeles, is valuable if it leads to lower wastage in an online advertising campaign. The costs of the advertising space and the targeting effectiveness determine its precise value.

If, for example, the goal is to reach only men but the gender of the recipient is unknown, it becomes statistically necessary to display two ads. But if gender is known in advance, one ad has the same effect. At costs of €1 (per thousand impressions) for a classic banner ad, savings of €1 (per thousand impressions) can be realized. In this case, the value of the information about gender amounts to €1 multiplied by the planned reach. If the information is used for video advertising costing €20 (per thousand contacts), the same information suddenly has a value of €20, twenty times the amount.

Data Quality is Hard to Quantify

The validity of the automatically generated user data is another fundamental problem. Verifying whether the cookies of online users describe them accurately is a service provided by third-party companies like Nielsen or Comscore. However, these companies use proprietary metrics, meaning that their measurements are not always consistent with one another. Tests in the USA and UK have shown that different validating companies assign different genders to one and the same online user. Some may categorize a user as male, while another categorizes the same person as female. As a result, even the data provider cannot be sure of the actual quality of the data in the lead up to an advertising campaign. The same applies to the data user. As long as there is no validation standard, the user cannot know which provider supplies good data, which means that they are taking a risk.

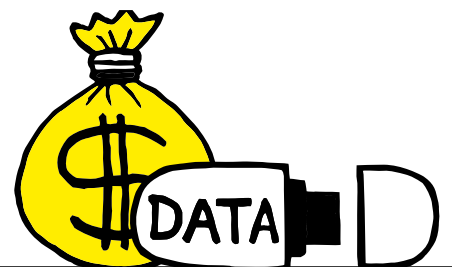
The following challenges are common in setting a price for data.

- > **A Suitable Price for Data is Hard to Determine** /// As a rule, the data provider will not know for what type of advertising its customers will use the data and is therefore unable to set an optimum price. As described in the box, the value of the data depends, for example, on whether they are used for display or video advertising. Accordingly, is it difficult to decide whether the data provider should demand a cost-per-thousand price of €1 or €20 for the "gender" characteristic? A possible solution would be a pricing model in which the provider participates in the cost savings of the data user. The price of the data could be set as x% of the costs for the saved advertising space. But because the data provider is not familiar with the costs of the advertising space and they are also not

»

Aside from data quality,
the biggest problem lies in determining
a reasonable price for the data.

«



known in advance for real-time bidding approaches, this pricing model has not been established in practice. Instead, billing is normally done according to a cost-per-thousand price. As a result, data used for display advertising is usually too expensive, while data for video advertising tends to be underpriced.

> ***The Customer Determines the Number of Contacts*** ///

Not only the price per impression is unclear. The number of contacts is also less obvious than it seems at first glance. In online advertising campaigns, it has become established practice for both the media seller (the publisher) and the media buyer (media agency, advertiser) to count how frequently advertising is displayed. Billing is done according to the cost-per-thousand price based on counts performed by both parties. Thus, if both parties count 1,000 delivered advertisements, the advertiser will pay the previously agreed price. Only in case of major discrepancies between the two measurements they would technically investigate the matter. To date, however, there is no standardized measurement of data usage for online advertisement. Measurements are usually taken by the advertiser, so that the data provider must rely on the accuracy of the information. Therefore the data provider is dependent on the data user's honesty.

> ***Price Markdowns for Questionable Data Quality*** ///

The problems with data validation described in the box can lead to diminished trust in the quality of the data to be purchased. It is thus not uncommon for buyers to demand risk-related markdowns due to the lack of assurance regarding the validity of the data. Whether such markdowns are justified is difficult to tell.

> ***Cost Structure Drives Price Pressure*** ///

When digital information has been collected once, e.g. that an online user is male, it can be sold as often as necessary. The selling costs are marginal and therefore the contribution margin is already positive, even at low prices. In a competitive environment, this situation leads to a downward price spiral. Data buyers can frequently negotiate significantly lower prices, because they know that the provider will still generate a positive contribution margin.

> ***Possible Data Theft*** ///

Another problem is that data users must incorporate the data into their own system in order to use it. Once incorporated, they can continue to use the data without paying for it. The data provider cannot

verify this or can do so only with considerable technical effort, which leads to additional dependency. The commercial success of the data provider is based on trust in the buyers and their proper use of the data.

From Wishful Dream to Nightmare to Reality? ///

The challenges described here have led many data providers to disappear from the market or be acquired by major technology companies like Oracle, Salesforce or Adobe. The use of big data, which is already a reality at Google, Facebook and other global players, has become a nightmare for them.

But are there solutions that could enable small data service providers to be successful? The most promising approach appears to be bundling data with advertising technology and advertising space. If these bundles come from one source and are sold as a combi-product, most of the problems involving price setting are eliminated. Big data can be better used to increase the effectiveness and inherent value of media, to achieve a margin from the cost savings of a highly targeted media selection and reduced wastage, and to leverage its economic potential completely.

Once the data quality problems are also solved, then the profitable use of big data is no longer a wishful dream for the data providers or their customers, but rather a reality.

./

FURTHER READING

<https://bigdatablog.de/big-data-3/>

<http://blogs.wsj.com/cmo/2015/02/23/marketers-question-quality-of-ad-targeting-data-providers/>