

DATA

ACCORDING TO WEBORAMA

APPLYING AUDIENCE
SEGMENTATION

TO MEDIA

MANAGEMENT USING

DATA

THE BIG DATA UNIVERSE

The development of digital technology and the new advertising opportunities it offers have totally transformed the advertising ecosystem which, since the golden age of Madison Avenue, was based primarily on creativity. Nevertheless we should not lose sight of the fact that data (now easily available thanks to the more recent media vehicles) “is not killing creativity, it’s just changing how we tell stories”.

(Alex Williams, journalist for *Tech Crunch*)*

S

ince the beginning of advertising, marketing professionals have dreamed of finding ways of talking only to useful audiences: in other words, to those potentially interested in their product or service. The world of communication understood the term “target markets” very early on: homemakers aged under 50, the upper middle-class, early adopters or seniors. The development of direct marketing and CRM or PRM databases helped to make targeting more accurate, making it easier to reach the consumer.

The internet and new technologies have completely changed the concept of the target market. The dream became a reality when it became possible to trace internet users’ web navigation. Advertisers can reach their clients and prospective clients by basing their operations on already completed by the internet user (buying, seeking advice, trying a product, etc.) as well as on their future needs, made more predictable by the monitoring and mathematical analysis of their online behaviour.

The Big Data phenomenon is expanding; brands can now centralise and exploit all sources of available data, whether online or offline, media or CRM. All with the view, on the one hand, of enriching their client knowledge with internet navigation data in order to build client loyalty, and, on the other, of creating databases of internet users who are potential purchasers of their products and services.

*“I think of that famous 1993 New Yorker cartoon by Peter Steiner: ‘On the Internet, nobody knows you’re a dog.’ Now it’s more like, ‘On the Internet, everybody knows what brand of dog food you buy.’”**

Big Data is not just about marketing, even if this sector clearly benefits from it. There is also finance, health, sport, and weather... according to journalist John Edwards, *“Like death and taxes, weather is a certainty. And thanks to Big Data, it’s now a more predictable certainty”*. Big

Data offers advantages in all areas, because it provides reliability, freshness and relevant data. However raw data is of no value. Adaptive technologies and human intelligence are required to give them meaning.

The difficulties related to Big Data concern the volume and heterogeneity of the data which must be processed extremely rapidly. Stephen Gold (IBM) estimated in 2012 that **“90% of all data in the world has been created in the past two years”**. But the great challenge for any data scientist is not just having unlimited storage or developing quick and powerful algorithms: data must be made relevant and dynamic; it has to tell us something. This is a more complex task.

For example: an algorithm is designed to detect where someone lives by analysing geolocalisation data linked to their mobile. The first results give three places: the person’s address, their office and... a red traffic light on a major road. Without human intelligence, processing Big Data cannot provide a relevant result.

To meet these challenges, Big Data draws on the latest technologies, as well as on scientific knowledge from the past. Developing statistical tests of a hypothesis or the first tree structure decisions - frequently used in segmentation marketing - date from the year 1900, and continue to be improved on today. On the other hand, “Map Reduce” is a modern data processing algorithm based on a method for calculating data distributed in parallel, very often seen in the world of Big Data, and first appearing in 2004. New storage models (the cloud, for example) and databases make relationship databases like SQL partially obsolete. The Big Data universe is continually moving.

Data experts have varied areas of competence, but what they all have in common is an openness to new technologies and the ability to apply previously acquired, and sometimes old, knowledge (theorems and algorithms) to these new technologies. Historical knowledge of mathematics and statistics is an advantage for any data expert.

ACCORDING TO DOUGLAS MERRILL, FOUNDER AND CEO OF ZESTFINANCE.COM, “With too little data, you won’t be able to make any conclusions that you trust. With loads of data you will find relationships that aren’t real... Big Data isn’t about bits, it’s about talent.”

* Geoff Nunberg teaches linguistics at the University of California in Berkeley. He is author of *The Way We Talk Now*, *Going Nuclear*, *Talking Right* and *The Years of Talking Dangerously*. His latest book released is *Ascent of the A-Word*. He is linguistics consultant in the NPR broadcast *Fresh Air*, hosted by Terry Gross.

WEBORAMA'S VISION OF “DATA SCIENCE”

DATA AT THE HEART OF WEBORAMA'S BUSINESS

Since it launched 15 years ago, **Weborama** has been involved in data science for online marketing.

Pioneers of technological innovation in their market, **Weborama** continuously invests in its R&D department, which employs 52% of all company employees. Data is at the heart of its expertise, thanks to its strategists, data consultants, computer linguists, statisticians, mathematicians and developers.

Big Data is the driver of development today at **Weborama** and will continue to be in the future. With the same care taken to collect data anonymously, the internet user benefits from research and studies built and calibrated on a mass of population.

RESPECT FOR PRIVACY

Since it was founded in 1998, **Weborama** has placed particular importance on respecting the privacy of internet users. The company plays an active role in a number of e-privacy projects in the digital sector in France.

A pioneer in this area, **Weborama** has been campaigning for several years for players in the market to behave with transparency towards consumers. For example, the company co-authored the white paper “targeted advertising and respect for the internet user” from the IAB. Other initiatives include the signing of the pan-European charter for behavioural advertising and the charter of the French Union of Direct and Digital Marketing (UFMD).

In accordance with current legislation, internet users can unsubscribe at any moment from Weborama's data collection programme, either by using the opt-out on the group's website, or via the icon Ad Choices, which can be found on all targeted advertising, and returning it via website YourOnlineChoices at IAB Europe.

The issue of internet user privacy was the subject of a book by Alain Levy, Chairman of **Weborama**, “In the footsteps of big brother, privacy in the digital age”*, extolling the importance of educating the wider public about digital. He concluded that: “[...] for internet users to accept this constant search for data, it is fundamentally important for companies in this area to be better known among the wider public. People must be kept better informed about ethical principles, and a platform of good practices must be consolidated. [...] They must engage in dialogue with their clients, employees, shareholders, political environment, associations, and consumers. [...]”. “This is not about policing anyone. We are just trying to improve our understanding of the tastes and habits of internet users so that they only see advertising messages which fit their profile and meet their expectations. It is the application of the principle of sustainable development to advertising.”

A SEMANTIC VISION OF THE INTERNET

The “data science” approach of **Weborama** consists of collecting and processing data, in order to analyse and segment this data so that it can be deployed in advertising campaigns. **Weborama** sees the web as an infinite lexical space where words are distant or close depending on the context in which they appear and how this context is revealed by searches by internet users and by the lexical content of the websites they visit when surfing. The data is sourced anonymously from 300,000 sites, using cookies, then processed instantly. This audience classification is a critical process within each database.

Thanks to this methodology **Weborama** has a database of 110 million profiles of behavioural, socio-demographic, and purchasing intent data in Europe, 70 million of which are from France.

* Alain Levy, “In the footsteps of big brother, privacy in the digital age”, L’Éditeur, 2010, Paris.

Today, **Weborama** has identified almost 200 clusters (centres of interest) based on the behaviour of internet users, divided into 23 segments (a segment being a “family” of clusters).

Weborama’s socio-demographic database is built from a sample group of 300,000 active internet users in France who responded willingly to a questionnaire giving their age, gender and socio-professional category. By cross-referencing behavioural clusters with **Weborama**’s socio-demographic sample group, the sample group data are extrapolated from the database as a whole.

USING TECHNOLOGY IN “DATA SCIENCE”

Weborama solutions as a whole are fed by patented technologies, based on their own algorithms and the automatic processing of language (*Traitement Automatique du Langage, TAL*). The robustness of the data collection and storage tools, and the solid infrastructure, allow instant processing of large volumes of data.

Thanks to intuitive navigation, the technological dimension moves aside to leave room for a data-centred user experience.

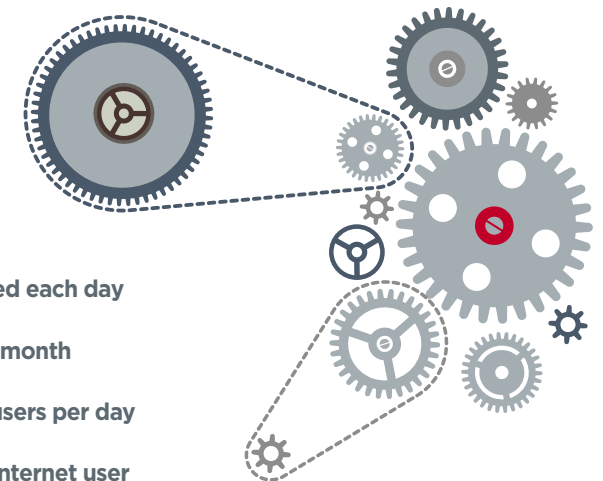
33 M active URLs

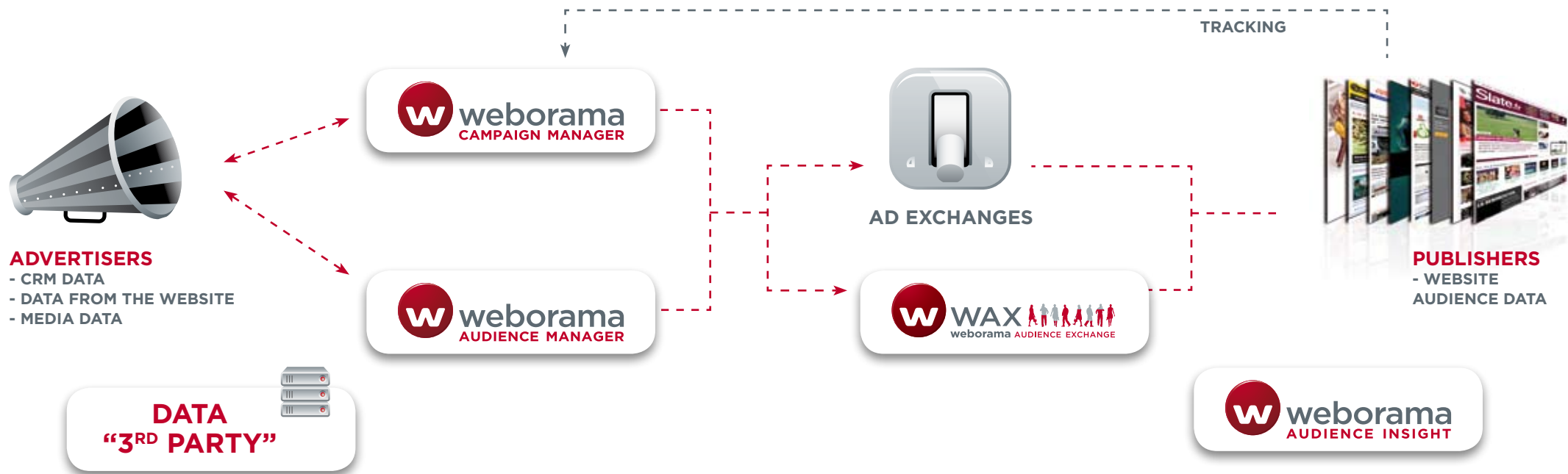
300,000 sites analysed each day

110 M active profiles per month

3.5 M qualified internet users per day

25 words per day and per internet user





DEPLOYMENT OF WEBORAMA KNOWLEDGE IN THE ADVERTISING MARKET

The 3rd party* Weborama database is a real source of 1st party** data enrichment, and a cornerstone of the set of tools available to clients. Weborama tested the efficiency of this data and has been a major user for 15 years. The data has also been used in branding and performance advertising campaigns (bundle data & media, offered by Weborama in its advertising platform, Weborama Audience exchange, and tested and compared to non-targeted campaigns) as well as in marketing research.

Weborama offers three technology solutions which complement and communicate with each other:

DATA SCIENCE

- Weborama Audience Manager (data management platform);

- Weborama Audience Insight (tool for analysing audience behaviour on internet sites);

TECHNOLOGY

- Weborama Campaign Manager (advertising tool for management and optimisation of advertising campaigns).

Thanks to these tools, Weborama has been successful in the transition from micro data to Big Data; in other words from a series of insignificant events completed by the internet user navigating the web, to the building of a complete marketing profile of this user.

* Data from a third party which do not belong to the advertiser or publisher, but are made available to them for deployment in online marketing.

** Data belonging to an advertiser or publisher.

POSSIBLE
APPLICATIONS OF
DATA
SCIENCE

& USE OF
3RD PARTY
DATA

SEGMENTATION FOR ADVERTISERS

THE INITIATIVE

There is a mine of information to be collected from the internet. For this to be of any use, a very large volume of available data must be processed and analysed, after determining which data will improve the ROI of the advertiser.

There are a number of techniques for targeting internet users. These techniques are often based on the advertiser's old techniques of target marketing and are not always as efficient as we might like. To establish which targeting of users is most accurate, we need to look at the performance of media campaigns, taking account of the internet user or "clickers" exposed to the campaign and the data relating to "transformations" (subscriptions to products offered in the campaign).

Segmentation of a website's audience or advertising campaign allows useful targeting of users to be determined according to the product or service offered by the advertiser. These target segments could be deployed as part of a future media campaign. With a view to extending the audience, it is now possible to target "look-alikes"; these are internet users with a similar profile to those who "transform". This allows them to be offered advertising which is personalised to fit their profile.

FIRST PRACTICAL APPLICATION BY AN ADVERTISER

Audience segmentation in media campaigns was first used in 2012. It was deployed by an advertiser whose catalogue was very developed, which made defining their target segments more complex. Although this player had a well-developed CRM, they were interested in learning more about potential target markets for each of their of-

fers and digitally available products, with a view to both winning over new clients and cross-selling.

The objective pursued was therefore the optimisation of digital campaigns to increase the level of transformations, while limiting loss of media investments. To achieve this objective, the advertiser called on **Weborama** to build a mechanism for its commercial catalogue.

WEBORAMA METHODOLOGY

Audience segmentation for advertisers consists of several stages. After the modelling phase, there is model validation phase, then a phase of tests within the media campaign. The initiative is integrated in the data process as a whole offered by **Weborama**: raw data collection, processing, analysis, segmentation and finally deployment in the media.

In the context of segmentation for internet management using data, the phases of the initiative can be translated as follows:

- creation of a database for data mining from one or several media campaigns carried out by the advertiser;
- model training, during which various assumptions are put forward on useful target markets for the products or services concerned;
- the calculation from segmentation trees;
- development of the targeted advertising model;
- simulation of a media campaign with the aim of confirming the assumptions put forward during the model training phase;
- dissemination of a targeted campaign and A/B testing (comparison of results between targeted and non-targeted population).

Creating a data mining database

Before segmenting the advertiser's audience, **Weborama** builds a database of internet users who have been exposed to their advertising campaigns, or visited their site. This database is then enriched with socio-demographic and behavioural criteria for internet users whose profile is known to **Weborama**, thanks to their database of 70 million qualified profiles in France. This data is cross-referenced with the data coming from **Weborama Campaign Manager**. This anonymous information is accessible as a whole using the internet user identifier (a series of characters for each individual contained in the cookies deposited in the internet explorer). This unique identifier allows **Weborama** to find the complete profile of the internet user in their database, in real time.

In this initiative, **Weborama** applied search data (from searches using search engines) and surf data (navigation through the web) for internet users, in order to assign them a profile relating to their gender, age and socio-professional category, as well as areas of interest and online behaviour.

For several weeks, **Weborama** gathered information on post view transformations (following advertising delivered by **Weborama**, which did not lead to a direct click at the time of viewing).

The data mining database built this way for the advertiser includes 60 million qualified profiles over approximately 300 columns, including: advertising to which the internet user has been exposed, their reaction to the advertising (visit, click), their behaviour on the website (categories of products searched for, products placed in the basket, conversion, etc.) and their socio-demographic and behavioural profile.

Training the model (calculation using the decision tree and development of targeted advertising)

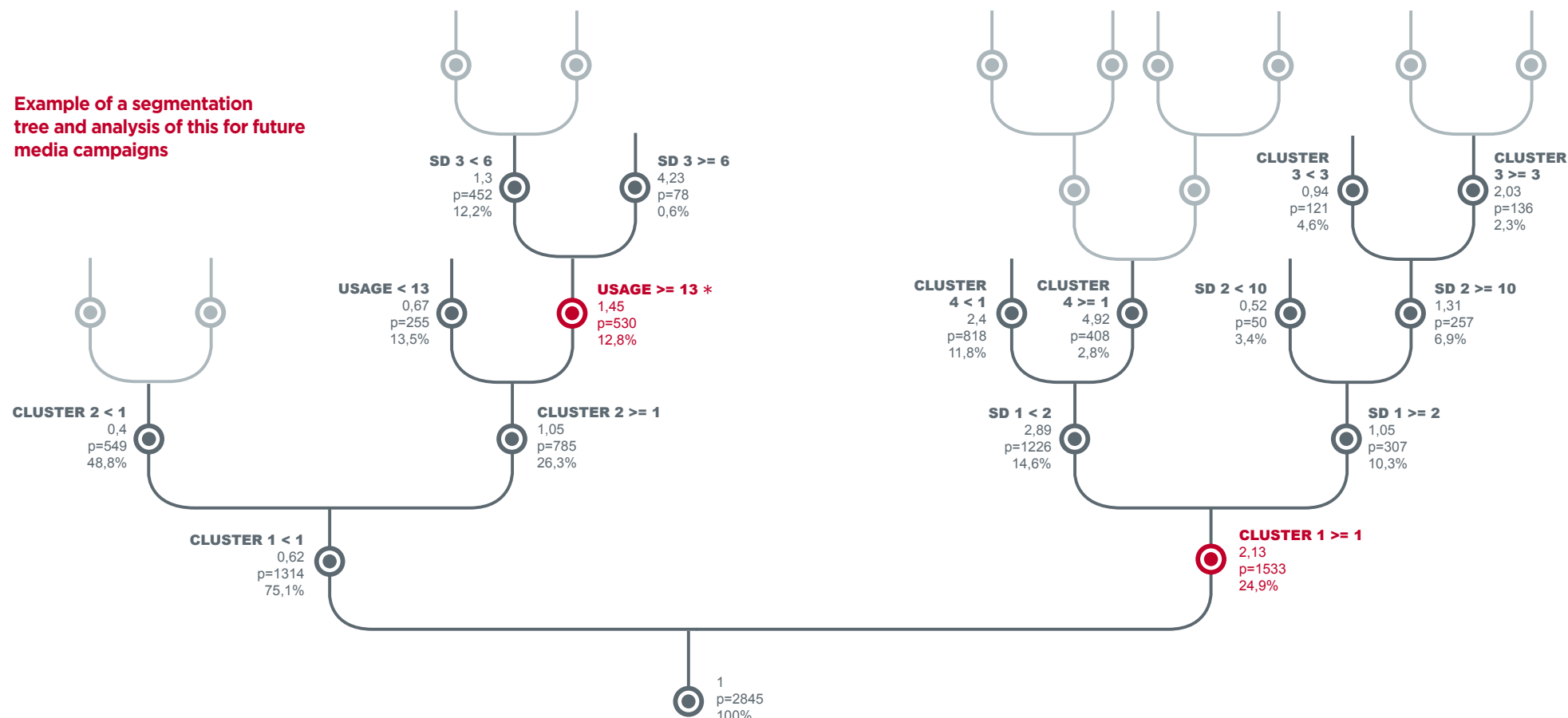
After creating the database used for segmentation, **Weborama** has built a graphic representation of the criteria for permanent target segments. Once the tree segmentation solution is created, the criteria most relevant to each

product or service on offer can be extracted. These criteria are based on the transformation of the internet user exposed to an advertising campaign and on the impressions created during the period of the campaign. The software chosen allowed each criterion to be tested, and a decision to be made on which criteria, contained in the cookies of the exposed internet user, to retain.

Before creating the segmentation trees, the **Weborama** segmentation algorithm allowed the most relevant criteria and affinity threshold to be assigned to each knot in the tree from the root up. **Weborama** then calculated the total gain in transformation levels for each knot in the trees. Only knots showing the highest levels of transformation gains will be selected for targeted media campaigns. **Weborama** also calculated a confidence interval for each gain in transformation levels. Between a minimum value and a maximum value, this interval will allow each model to be confirmed in the media campaign simulation phase, as well as allowing the results of the advertiser's future targeted campaigns to be monitored (*see below*).

The advertiser's proposal is complex and includes different variants (5 products in total). Audience segmentation will therefore be completed taking account of these characteristics, with a segmentation tree for each product.

Example of a segmentation tree and analysis of this for future media campaigns



The segmentation tree is read from bottom to top, from the root to the upper branches. At the root, the whole population exposed to the campaign is represented, with the gain in level of transformations equal to 1, for the purpose of comparison with the other levels of branches, the number of positive examples (internet users who changed) and the percentage of advertising imprints spread among the population (here equal to 100%).

The branches of the tree contain 4 key pieces of information:

- the behavioural cluster or determining socio-demographic criteria, with the relevant quartile;

- the gain in level of transformations compared to the global exposed population. This is only relevant to targeting in a media campaign if it is over 1. If the reverse is true, it can be used to exclude some communication mechanisms for internet users online who do not transform;
- the number of positive examples in the population belonging to a criteria or a cluster;
- the percentage of advertising imprints spread among the population. This will be an approximate percentage of the population exposed to the campaign.

The analysis of the segmentation tree rests on the ability to find a balance between performance - in other words, useful target segments (a gain in transformation levels higher than 1) - and the percentage of imprints sufficient for the strength of the campaign.

* "Usage" criteria relates to internet users whose internet navigation is supported.

Two branches from the same level always complement each other. Internet users cannot be included on both sides at the same time. Several internet user groups can therefore be added (*see red criteria in the example above*) to create a broader distribution.

Example of reading: let us focus on the two populations highlighted in red in the decision tree. On the one hand we have internet users who, in their cookie, have CLUSTER 1, with a quartile above or equal to 1. We know that they have been exposed to the campaign and have transformed their behaviour 2.13 times more than the overall population. We also know that this represents almost 25% of the exposed population.

On the opposite branch, we are interested in internet users who have, **SIMULTANEOUSLY**, cluster 1, with a quartile less than 1 (therefore equal to 0) **AND** cluster 2, with a quartile equal to, or higher than, 1, **AND** who are regular surfers (this criteria must reach the minimum quartile of 13). These internet users make up almost 13% of the population exposed and transform their behaviour 1.45 times more than internet users who potentially saw the campaign.

Simulation of a media campaign (confirmation of assumptions)

The definition of an audience model for each product offered is just a modelling phase. It must be confirmed by simulation of a media campaign. No targeting is introduced. The results of this campaign are cross-referenced with the data mining record, to see the correlation between the model previously defined during model training.

In the case of simulation for the advertiser mentioned above **Weborama** observed perfect connections between the data mining database and the simulation of a campaign. These results from application of the segmentation trees have allowed the accuracy of the proposed targeting to be confirmed for each product offered. The profiles which transform are in fact those which **Weborama** has put forward in the decision trees. Future media communication can therefore benefit from more accurate targeting to increase the digital ROI for each campaign.

Dissemination of a targeted campaign and AB testing

After defining the targeting models assigned to each product offered and following confirmation from the decision trees during a simulated campaign, the data will play its final role. It will then be assessed within the media campaign in order to improve efficiency of the data and reduce loss.

Simultaneously, and in exactly the same conditions, Weborama launched two media campaigns: one targeted campaign and another to a non-targeted population. The comparison of the results of the two types of campaigns once again confirmed the accuracy of the specified target segments defined above in the decision trees. An increase in the number of the expected transformations was indeed achieved.

Variations in gains in transformation levels were noted with regard to those exposed in the decision tree. Nevertheless, the results did not exceed the confidence interval.

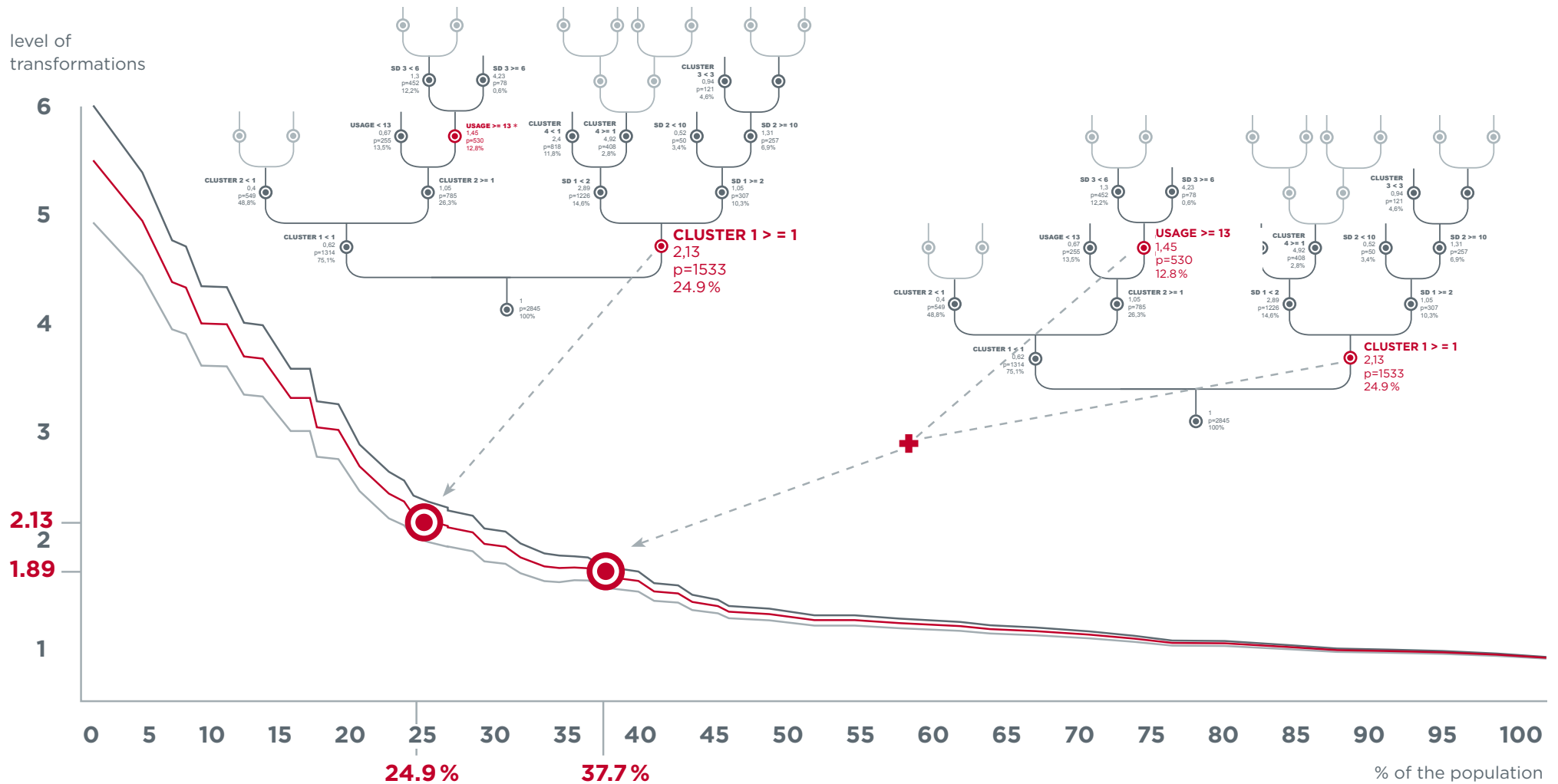
The variation in results compared to the forecasts may have occurred for a number of reasons: they may be related to the creative element of the advertising or to factors external to the campaign, like a calendar event, or a holiday period, in which internet users change their web navigation habits. The advantage of designing confidence intervals in advance is based on all these factors.

Deployment and future of audience segmentation

Audience segmentation, and specifically segmentation trees, can be used for different purposes. In the context of internet management using data, beyond improving the digital ROI of a media campaign by targeting internet users potentially interested in their products, it is possible to apply the trees to cross-selling or calculating the potential value of an internet user.

Weborama is developing the integration of this mechanism into its tools for marketing control using data: C-clones. C-clones, or “look-alikes” are an extension of the audience, based on the transformation of internet users’ behaviour. As in the initiative described above, the performance and strength of an effective media campaign

must be enhanced by selecting a population which is likely to subscribe to products in the campaign, and which is large enough for dissemination. The C-clones interface will be integrated in the data management platform **Weborama Audience Manager**. By integrating segmentation trees, the useful profile of the campaign can be determined at a glance.



SEGMENTATION FOR PUBLISHERS

THE INITIATIVE

Knowing their audience has become a major challenge for publishers, who want to get the best from their advertising to optimise commercial value. Segmentation of a website's audience offers the publisher a number of opportunities.

OPERATIONAL OPPORTUNITIES

Above all, the aim of publishers is to organise their commercial products so that they can offer advertisers and agencies, in addition to the contextual menu, packs of targeted audiences, enriched by 3rd party data. The second challenge is to build specific audience packs using ad hoc target segments based on both socio-demographic and behavioural targeting (taking account of the internet users' different areas of interest). Exploitation of the data will allow the commercialisation of the site's product inventory to be managed and advertising to be disseminated to useful contacts, while at the same time "unclogging" overloaded headings. The cost per thousand sales of advertising space will therefore be reassessed according to sales per useful audience.

So-called event advertising formats are mainly commercialised by subscription, on a daily basis. Using the data, they can also be targeted and made available to buyers according to CPM. To give an example, "behavioural takeover" will be visible for more than a day, and not just on the home page. It will be visible for a period of 15 days over all the pages on the site and posted according to the internet user's profile. It will be relevant to the target user, and related to the campaign.

Segmentation of a site's audience allows visitor profiles from other areas of interest to be seen. It is highly probable that a current affairs site has, among its audience, people who enjoy other sites (finance, beauty, fashion, cookery, travel...). This will offer the publishers the possibility of assessing their content quality, as well as that of their audience, while expanding their client base.

COMMITMENT

Weborama cooperates with publishers to collect raw and anonymous navigation data. Audience analysis, followed by a relevant and targeted marketing proposal will be conducted within three weeks from the date the Weborama tags are placed on the publisher's site. The data to be used in the target segmentation will be transferred within a month and the target users will be optimised daily, in order to best respond to the expectations of the publisher.

In the same way, Weborama will not commercialise the data alone and will not re-define the site's audience packs. Information linked to the audience of the site will be strictly confidential. Weborama forbids any communication of this data to other players.

WEBORAMA METHODOLOGY

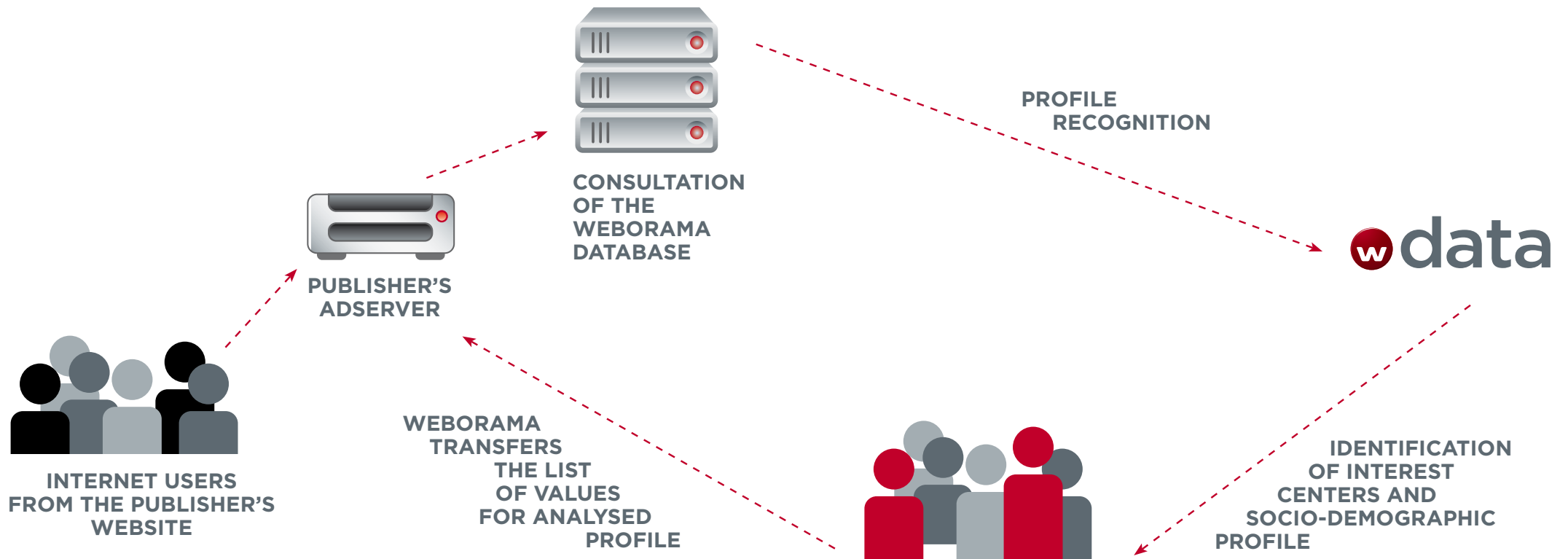
Weborama offers two methods of data enrichment to publishers using 3rd party data.

The first method includes the synchronisation and propagation of **Weborama**'s socio-demographic criteria and behavioural clusters on the publisher's site audience. This is possible by identifying profiles recognised within the site's audience, which are already part of the **Weborama** database.

Synchronisation and propagation of clusters and socio-demographics

In this particular case, the adserver of the publisher calls the **Weborama** database for each internet user. If the **Weborama** database recognises the socio-demographic profile or behaviour of the visitor to the site, the visitor is transferred to the publisher's adserver.

The publisher's adserver therefore now contains the socio-demographic and behavioural profile of the majority of its internet users.



The second method identifies and analyses the website user's profiles to create *ad hoc* target users relevant to the advertisers' areas of activity.

Identification, analysis and creation of *ad hoc* targets

This enrichment is geared towards the analysis of profiles of users of the publisher's website to create *ad hoc* target users which are their own, and can be used by media agency.

There are three phases in this approach: firstly, the target users are identified and analysed, then the target segments are enriched and customised target users created for the media agency, and finally the profile transfer process takes place.

- Identification and analysis of target users

In this stage, **Weborama** introduces a tool for the behavioural and socio-demographic analysis of the site, **Weborama Audience Insight (WAI)**. WAI allows beha-

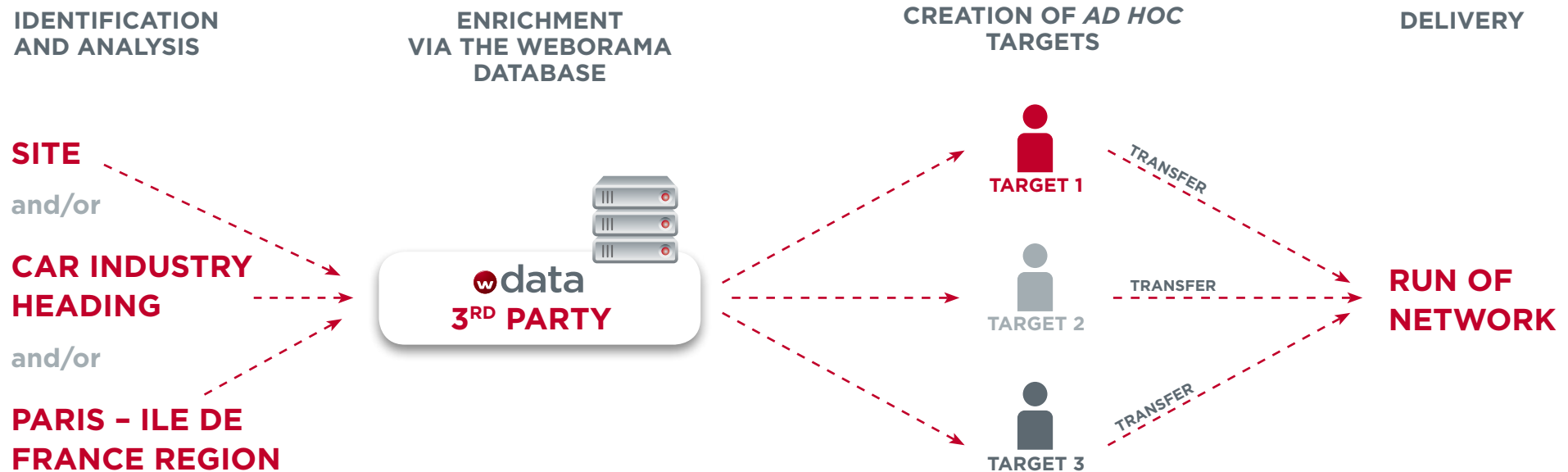
vioural segments and clusters to be identified by strength and/or affinity, by cross-referencing the site audience with the **Weborama** database. The most discriminating profiles are deduced by comparing the audience of the publisher's site with the average audience taken from the **Weborama** database.

- Enriching the target segmentation and creating *ad hoc* target users.

The **WAI** study allows *ad hoc* target users to be created by linking socio-demographic criteria with behavioural clusters. Finding a balance between indicators of reach and affinity determines the volume of relevant profiles.

- Transfer of *ad hoc* target users.

The transfer of *ad hoc* users defined above is possible using a list of different adserver pixels. This transfer is done by a call of the pixels generated by the publisher's adserver. The volume of profiles per target required are then available directly in the adserver.



DATA DRIVEN DIGITAL MARKETING

D

ata will not just affect actions related to digital communication. It is increasingly present and will play an important role in all aspects of our lives, both personally and professionally. In order to take advantage of the opportunities provided by “the gold of tomorrow”, companies must review their internal organisation, often structured around compartments such as: digital media, commercial website, CRM, etc.

In the advertising market, data becomes a completely separate entity from marketing. Today it has to be deployed in all areas of digital brand marketing. These brands need to find ways of reaching clients, and prospective clients, in a more relevant way so that their online communication is more effective. In parallel, the dichotomy of offline and online is disappearing.

For the last 15 years, **Weborama** has been offering innovative and complete solutions. Audience segmentation and decision trees are just one part of **Weborama**'s data science. In an effort to make this science based on complex algorithms, storage technologies and robust processing more easily accessible to users, **Weborama** will continue with its research and development. The passion for data is in the DNA of the company. Well-known in the European advertising market its lexical approach to digital marketing, the claim “from data to value” is a clear illustration of their knowledge and willingness to make data useful by adding the human dimension.

WEBORAMA, FROM DATA TO VALUE

Weborama is the European market leader in data, with a database of 110 million profiles.

Weborama has been developing solutions in adserving, Data Management Platform (DMP), audience analysis, data segmentation and target segments for 15 years.

Over 300 clients use **Weborama's** innovative solutions to drive and optimise their online investments in France, Spain, Italy, Portugal, Holland, UK and Russia:

Weborama Audience Manager: data management platform for segmentation of different sources of 1st party (CRM, media and website) data, enriched by 3rd party data using access to data owned by **Weborama**

Weborama Audience Insight: a tool for behavioural and socio-demographic analysis of internet sites

Weborama Campaign Manager: adserving and ad analytics for management and optimisation of advertising campaigns.

Weborama Advanced Rich Media: tool for managing rich media/video campaigns.

Weborama Audience Exchange: premium multi-device audience platform, on two channels (WAX for agency, WAX for trading desk). Classic formats, rich media and in stream.

Weborama *
has been listed
on Alternext
since June 2006
(ALWeB).
It recorded
revenues
of 24.3 million
Euros in 2012.

© 2013, Weborama.

All rights reserved for all countries.

Any reproduction, even partial, is strictly forbidden.

Printed in France.

Graphic design © LES 5 SUR 5 .COM



www.weborama.com