MULTISENSOR

Mining and Understanding of multilinguaLcontenT for Intelligent Sentiment Enriched coNtext and Social Oriented inteRpretation

FP7-610411

D3.2

Indicators for media monitoring and internationalisation

Dissemination level:	Public
Contractual date of delivery:	Month 16, 28/2/2015
Actual date of delivery:	Month 16, 28/2/2015
Work package:	WP3 User and context-centric content analysis
Task:	T3.1 Indicators for media monitoring and internationalisation
Type:	Report
Approval Status:	Final Draft
Version:	1.0
Number of pages:	56
Filename:	D3.2_Indicators_2015-02-27_v1.0.pdf

Abstract

The objective of this document is to report on indicators for media monitoring, journalism and SME internationalisation. These indicators are defined both as conceptual factors and measurable metrics that could be used in order to perform reasoning and derive high level conclusions in journalistic, media monitoring and SME internationalisation use cases.

Co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	24/06/2014	Initial structure	M. Eckhoff (PR)
0.2	15/07/2014	Final structure	ALL
0.3	01/10/2014	Input by PR	M. Eckhoff, F. Klaus and M. Menken (PR)
0.4	14/01/2015	Input by DW	N. Heise and T. Wagner (DW)
0.5	02/02/2015	Input by BM-Y	I. Arapakis and I. Miliaraki (BM)
0.6	24/02/2015	Final Draft for internal review	T. Forrellat (PIMEC)
0.7	25/02/2015	Final Draft review	A. Mas (everis)
1.0	26/02/2015	Final version incorporating the review comments	T. Forrellat (PIMEC)

Author list

Organisation	Name	Contact Information
PR	Mirja Eckhoff	mirja.eckhoff@pressrelations.de
PR	Michaela Menken	michaela.menken@pressrelations.de
PR	Florian Klaus	florian.klaus@pressrelations.de
PIMEC	Teresa Forrellat	tforrellat@pimec.org
DW	Nico Heise	nicolaus.heise@dw.de
DW	Tilman Wagner	tilman.wagner@dw.de
BM-Y!	Iris Miliaraki	irismili@yahoo-inc.com
BM-Y!	Ioannis Arapakis	arapakis@yahoo-inc.com

Executive Summary

This deliverable presents and describes a number of indicators useful for the tasks of the different use cases in MULTISENSOR including commercial media monitoring, journalism and SME internationalisation. Overall, the main contributions of this deliverable are a) the definition and the analysis of the indicators from the user perspective and b) the attempt to quantify these indicaotors and identify ways to extract and calculate them.

Depending on each use case and the nature of the task that is being tackled, the indicators are either high-level (conceptual) or low-level measurable units that can be computed or extracted by tools developed in MULTISENSOR or accessible APIs. In many cases during the process of the indicator definition we start from the conceptual indicators, which we tried to translate into measurable metrics. Towards the definition of measurable indicators we investigate the application of specific mathematical models. In this case we have mostly focused in deriving such metrics for indicators that are relevant to specific research activities of the project such as social media analysis. Finally, we provide a detailed summary of the indicators for each use case and how each of them can be extracted and/or calculated. In the next phase of the project, these indicators will be used by the reasoning and the decision support modules especially for assisting the decision making process in the SME internationalisation scenario. The indicators for media monitoring and journalism will also be considered for the derivation of important facts using the reasoning services.

Abbreviations and Acronyms

AGMA	Arbeitsgemeinschaft Media-Analyse
AMEC	International association for the measurement and evaluation of communication
API	Application Programming Interface
AVE	Advertising Value Equivalency
AWA	AllensbachWerbeträgerAnalyse
CPE	Corruption perception Index
DW	Deutsche Welle
EBU	European Broadcast Union
GDP	Gross domestic product
HDI	Human Development Index
PEST	Political, Economic, Social and Technological
PESTL	Political, Economic, Social, Technological and Legal
PPP	Purchasing Power Parity
PR	Pressrelations
RTV	Radio Television
SME	Small and medium-sized enterprises

Table of Contents

1 INTRODUCTION

One of the important objectives of MULTISENSOR is to deliver intelligent reasoning services in the context of three use cases: a) journalism, b) media monitoring and c) SME internationalisation. Especially for the SME internationalisation a decision support mechanism is foreseen to advice investors and entrepreneurs during the procedure of internationalisation. In order to provide such high level conclusions the reasoning and decision support services (WP5) have to be based on specific criteria that are combined in a meaningful way. This deliverable aims at defining these criteria as conceptual and measurable indicators. D3.2 describes the work done in Task 3.1 (WP3) during the first 16 months of the project.

This deliverable is organised as follows:

- **Section 2**: **Definition and methodology**. Here we define what an indicator is and the methodology followed to discover and define useful indicators.
- **Section 3**: **Indicators in journalism context.** In ths section we describe the indicators in the journalism context.
- **Section 4**: **Indicators in the commercial media monitoring context.** We describe the indicators in the commercial media monitoring context and we provide examples and formulas for indicators that are measurable and can be computed.
- **Section 5**: **Indicators relevant for SME internationalisation**. In section 5 we describe and give some examples for the important indicators for SME.
- **Section 6**: **Extraction of indicators in MULTISENSOR**. We explain how we can the proposed indicators could be quantified.
- **Section 7: Calculation and summary of indicators**. We include here three tables, one per use case, which summarise all the indicators along with a description, a method of computation or extraction, and finally information regarding data acquisition, pointing to relevant sources, APIs or modules that can be used for extraction.
- **Section 8**: **Conclusions**. We summarise the work presented so far and present conclusions and findings.

2 INDICATOR DEFINITION AND METHODOLOGY

In MULTISENSOR, an **indicator** is defined as a measurable quantity (e.g., a number or a ratio) or any piece of conceptual information (i.e., non-measurable) derived from a series of observed facts serving as a guide to the process of **decision making.**

Based on the above definition and considering the diversity within the set of indicators considered by each use case, we categorize the indicators in two different types, depending mainly on whether they refer to something measurable or not:

- Conceptual Indicators: Such indicators refer to high-level concepts, such as background information for a specific media source or a timeline of events connected to an emerging topic. These indicators cannot be directly measured but usually require the involvement and an analysis by the end user.
- Measurable Indicators: Such indicators are computable, measurable units derived by collections of data; such examples include the GDP of a country, the number of followers of a Twitter user, the number of articles for a specific topic or the circulation of a specific media.

For each use case we have followed a mixed approach of gathering indicators for achieving the different analytical tasks. The exact measures depended on the use case. The analysis starts from the conceptual indicators, which we try to model as measurable. As it will be presented also in section 7, several conceptual indicators cannot be easily translated into measurable indicators.

The indicators provided for the journalism use case serve in the process of journalistic research by considering three different steps; (1) Identification of topics and relevance check, (2) verification, and (3) completion and enrichment. These basic steps in journalistic research require different indicators to achieve a better understanding of the background of essential journalistic task. For example, for discovering a new topic, a useful indicator is the list of trending topics provided by Twitter or in the context of verifying an occurred event; a useful indicator is the proximity of the source to location of the event. The descriptions of indicators in the framework of journalistic research are based on common (journalistic) knowledge, as well as extensive discussions with journalists in DW.

For the media monitoring use case, we divided the indicators into two groups; on one hand the indicators in traditional media and on the other hand the indicators in social media. This way, we separate the traditional indicators that already been used with the new ones that started gaining ground during the last couple of years. These indicators are needed to generate different types of insights within media monitoring campaign carried out by a company such as PR. The set of provided indicators has been collected by interviewing the internal personnel of PR and refer to metrics currently exploited in the internal processes of the company. These processes are also guided by the various requests of numerous PR clients for media monitoring.

Furthermore, the indicators for SMEs are required in order to prioritize the market and strongly depend on the SME that is interested in exporting given the fact that not every company has the same economic capacity or production capacity. In this case we define the indicators based on the steps of the PESTL Analysis (Political, Economic, Social, Technological

and Legal). In order to decide to use PESTL Analysis for this purpose, we asked several export managers in PIMEC inquiring also about which kind of indicators could cover different types of companies.

In the following, we describe all the indicators focusing on the view of the final user and how they assess their needs. At the end, we provided also a technical view discussing how each indicator will be computed, extracted or otherwise assessed pointing to the relevant tools or available sources by considering also the input by the empirical study deliverable (D2.1).

3 JOURNALISM INDICATORS

In this section, we describe the standard approaches and indicators that are used by journalists with regard to research. In order to better understand the background of this essential journalistic task, we will briefly describe the basic steps and rules of professional journalistic research.

The goal of any journalistic research is to analyse and understand a piece of information or an event as accurately and comprehensively as possible, to put it into context and to publish the results to the audience. In order to achieve this goal, journalistic research usually follows *three* basic steps:

1. *Identification of topics and relevance check*

   The first step in the journalistic research work cycle is to identify a new topic and to assess its relevance. In case the journalist is already working on a specific topic, he/she needs to establish whether any new piece of information that is identified is relevant to this topic.

2. *Verification*

   Secondly, a new piece of information needs to be verified. The journalistic code of conduct applies strict rules with regard to verifying information before publishing it. During this verification process, the journalist checks the validity of the information itself as well as the trustworthiness of its source.

3. *Completion and enrichment*

   Finally, a journalist looks for enriching the (verified) information that he has already collected by expanding his research to further related content and information. The ultimate goal is to achieve a complete and accurate picture of the chosen topic.

These three main steps in journalistic research do not necessarily need to be always followed in the same order. They are sometimes intertwined and complement each other. For instance, any new piece of information that a journalist has identified and that he/she wants to add to his/her story needs to be checked for its relevance and validity again.

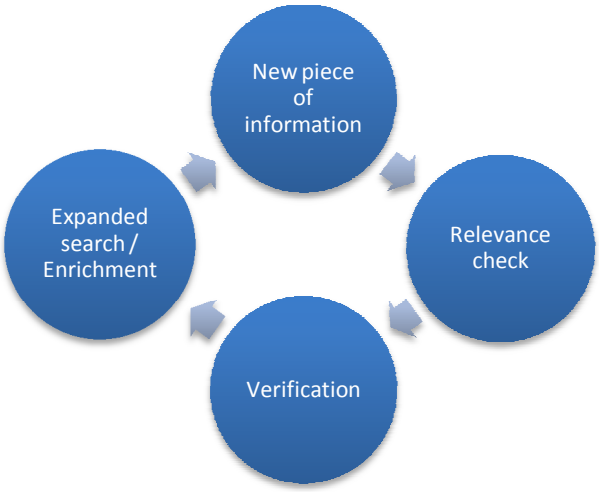


Figure 1: This is the journalistic research cycle

3.1 Identification of topics and relevance check

In order to identify a new topic or a piece of information that is relevant to an already followed topic, a journalist observes the general media coverage as well as other sources. This could be (amongst others):

- Topics that are already covered by other media outlets
- Information from news agencies and professional news correspondents
- Trends that derive from the internet as a whole and especially from social networks

In order to identify a new topic, a journalist might check how frequently a specific subject is mentioned by different sources. He/she might also look at trending topics on social networks, by following trending hashtags, the number of posts, tweets or likes towards a specific topic as well as other patterns and anomalies.

After having identified a new topic (or a new piece of information), the journalist assesses its relevance by asking the following questions:

- Is this topic / piece of information of general public interest?
- Is this topic / piece of information of interest for my specific target group?
- Is this topic / piece of information of interest for me / my specific media outlet?

Apart from checking the content itself, relevance could also be indicated by how a topic is discussed in the general public or very specific interest groups. The following overview (Table 1) shows possible indicators for this stage in the journalistic research cycle:

Research task	Possible indicators
Identification of topics / information	• Number of mentions by other media outlets • Number of mentions by news agencies / independent correspondents • Number of mentions by main influential with regard to a specific topic (so-called „news hounds“) • Number of related hashtags • Position in the trending topics lists • Number of tweets/retweets, likes, shares etc. • Patterns or anomalies (e.g. in big data analysis)
Relevance check	• The content / information itself • Mention of same specific keywords, named entities, urls etc. • Public discussion (*extended public discussion might imply relevance for the general public*) • Related traffic (*heavy traffic might imply high relevance*)

Table 1: Indicators for journalistic research cycle

3.2 Verification

Verification is the most crucial task for any journalist. It is part of the journalistic self-understanding and underlying principle of the journalistic code of conduct. In general, this means that no piece of information can be published if it has not been properly verified. This includes *fact checking* - the examination of the validity of the underlying facts - as well as the assessment of *how reliable the source* is.

3.2.1 Fact check

Verification always starts with checking the underlying facts. In doing this, journalists usually focus on the following six questions:

***Who** acted? - **What** happened? - **When** did it happen? - **Where** did it happen?*

*- **How** did it happen? - **Why** did it happen?*

1. ***Who** acted?*
   Journalists need to confirm the identity of those individuals who caused or participated in a specific event.
2. ***What** happened?*
   Journalists need to check what exactly happened. Who did what and what was the outcome?
3. ***When** did it happen?*
   Journalists need to check at which time the event happened.
4. ***Where** did it happen?*
   Journalists need to confirm the location of the event.

The answers to these first four questions should be rather clear, as the number of possible answers is limited. The answers to the remaining two questions

5. ***How** did it happen?*
6. ***Why** did it happen?*

On the other hand are rather ambiguous. They are based on the interpretation of individual facts and therefore very much depend on the journalist's analytical skills. The journalist needs to identify the process behind a specific event. This includes the underlying chronology but also needs to consider a lot of additional information such as the motives of the involved individuals, their influence and their room for manoeuvre.

The following overview (Table 2) shows indicators that are relevant to journalists when checking facts:

Research task	Indicators
Who?	• Names • URLs • Profiles (social networks) • Location • Images
What?	• Textual description • Images & videos • Image and video analysis (*is the content genuine or has it been manipulated?*) • Reports by others (police and other state authorities, witnesses, media outlets) • Duplication (*multiple reports about the same event)*
When?	• Time stamp • Context (*was it dark? did the sun shine?*) • Content metadata
Where?	• Geo location • Implied location • Content metadata
How?	• Timeline (chronology) • Images/videos • Reports by others (police and other state authorities, witnesses, media outlets)
Why?	• History • Background and position of involved individuals • Interests / motives • Influencer scores (e.g. Klout) and popularity • Activity (e.g. in social networks) • Money flow • Sentiment and tonality • Assessment by others

Table 2: Indicators that are relevant to journalists when checking facts

3.2.2 **Reliability of the source**

When information comes from a specific source, a journalist needs to assess its reliability. Some indicators could be the source's specific knowledge about the topic in question, its

trustworthiness in the past, its influence in a specific network, its reputation, the number of followers, likes and retweets or the genuineness of its profile or the URL in use. Another fundamental role of journalistic research is that a source is considered as more reliable the more impartial it appears with regard to the event it reported. Consequently, a source might be considered as less reliable if it has personal motives with regard to the reported event, when it is involved with other individuals that play a role in context with the topic or when personal political and/or financial interests are at stake. An overview of relevant indicators is provided in Table 3.

Research task	Indicators
Source reliability	• History of source / url / profile • Geographic proximity of the source • Specific knowledge of the source • Political / financial background • Influencer scores (e.g. Klout) and popularity • Position in network (network visualisation) • Activity (e.g. in social networks) • Number of followers/retweets/likes/shares • Sentiment and tonality

Table 3: Indicators that are relevant to journalists when considering source reliability

3.3 Completion and enrichment

After having identified a relevant topic or a relevant piece of information and after having verified it, a journalist usually strives for completing the picture. One reason for this is that one piece of information (as true as it might be) is often not sufficient to tell a story. It needs additional information and content in order to understand different angles, to come to valid conclusions and also to present the story in an appealing way that is easy to digest for the audience.

In order to do so, a journalist looks for additional sources and information. He searches for similar or contradicting quotes. He/she tries to find related images/videos as well as data and information that lead to deeper understanding. He/she also analyses the background of a topic or event as well as of the individuals involved in the case in order to bring out underlying links and interests.

Furthermore, the journalist will try to find additional content in order to enrich the story and to make it more attractive and compelling for his/her audience.

The following overview (Table 4) shows indicators that are relevant for the completion and enrichment task.

Research task	Possible indicators
Completion	• The content / information itself • Mention of same specific keywords, named entities, URLs etc. • Public discussion • Related traffic
Enrichment	• Mention of same or related keywords • Image similarity • Patterns/anomalies (e.g. in big data analysis) • Popularity (of content) • Context analysis

Table 4: Indicators that are relevant to journalists for the completion and enrichment task

4 INDICATORS IN COMMERCIAL MEDIA MONITORING

In this section we describe the standard approaches and indicators used in media resonance analysis projects. This analysis takes into account different analysis formats (presence analysis, benchmark analysis, readership structure analysis, stakeholder analysis, issues analysis, event analysis, reputation analysis, and AVE analysis) and present existing industry standards. Additionally, less common indicators will be discussed and evaluated.

Media monitoring is basically a structured evaluation of internal and external information, gathered from a collection of data. This information is stored and analysed. Depending on the client's focus, different indicators are monitored and taken into account for the analysis. Some indicators can be considered central, e.g., every client is interested in the absolute number of news or media items that have been found for one specific search profile. Also, the amount of people that have been ultimately reached by one particular article is generally relevant information. On the other hand, not everybody is interested in social media monitoring, although we can see a clear trend that social media gains more and more acceptance as an opinion outlet that needs to be taken seriously.

We divide the indicators for commercial media monitoring into two groups: the indicators in traditional media on the one hand, and the indicators in social media on the other. Traditional media not only covers print media and radio/television, it also subsumes online publications of all sorts. At the moment, pressrelations monitors around 32000 online media, 27000 of which cover international online media publishing news-like content, and approximately 5000 social media sources are monitored. The indicators that are taken into account for the set of traditional media are fairly well described, and there is only little disagreement between different media monitors on how to use and correlate the different indicators. The second group of indicators is far less standardised than the first, which is directly related to the observation that social media monitoring only started gaining ground during the last couple of years. But while even very traditional companies begin to see great additional value in analyses of social media information, particularly for purposes of issue management, the monitoring industry is still far away from a homogeneous set of indicators and correlations that need to be taken into account when analysing the social media resonance for a client. In section 4.3, we will discuss the social media indicators that are monitored at pressrelations to this date, plus additional ones that will prove useful but are difficult to track with the methods we are using so far.

4.1 Types of analysis in Media Monitoring

Different clients want to generate different types of insights with their media monitoring campaign. And since no two campaigns are alike, there can be no single description of a default commercial media analysis. In this section, we will describe the different types of campaigns that are offered by media monitoring companies,[1] followed by an in-depth analysis of the different indicators that can be measured or calculated from a given set of news items.

[1]All information about different types of media analysis campaigns stems from pressrelations' internal material, as well as http://www.pressrelations.com/pressrelations/index.cfm/en/analytical-approach.

4.1.1 Benchmark analysis

The aim of a benchmark analysis is to determine and evaluate the positioning of a client in the media as opposed to its key competitors. A benchmark analysis allows for comparative statements across all possible levels of analysis. Whatever indicator is measured, it is compared to the results for a predefined set of competitors.

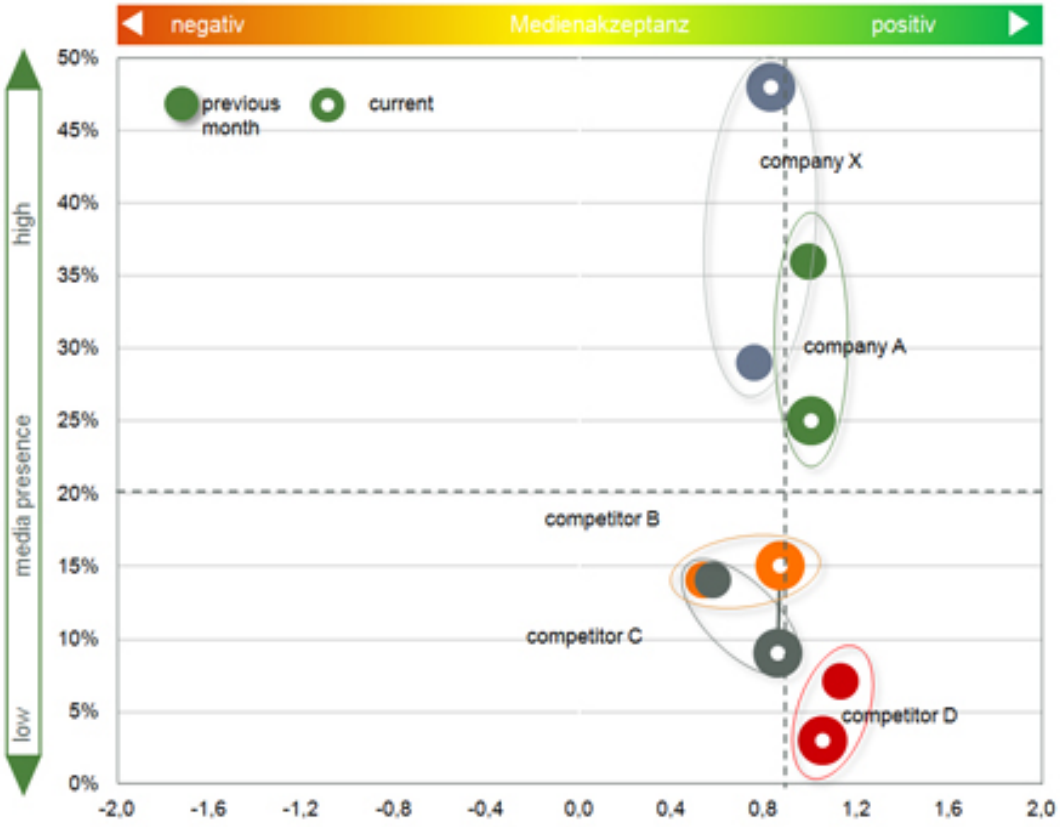


Figure 2: This is the benchmark analysis

4.1.2 Event analysis

An event analysis focuses on the media response to a particular event, possibly a press conference, an official statement or an extraordinary happening like the opening of a new branch, etc. The goal of an event analysis is to provide information about the cross medial impact of the event and to evaluate medial feedback quantitatively and qualitatively.

Figure 3: This is the event analysis

4.1.3 **Presence analysis**

In a presence analysis, only structural criteria of an article (circulation, type of media, etc.) are taken into consideration, while no evaluation of the editorial content is made. Yet, a purely quantitative analysis of media coverage can provide general information on when, how often, in which media source, and with which coverage the company or product has been reported on.

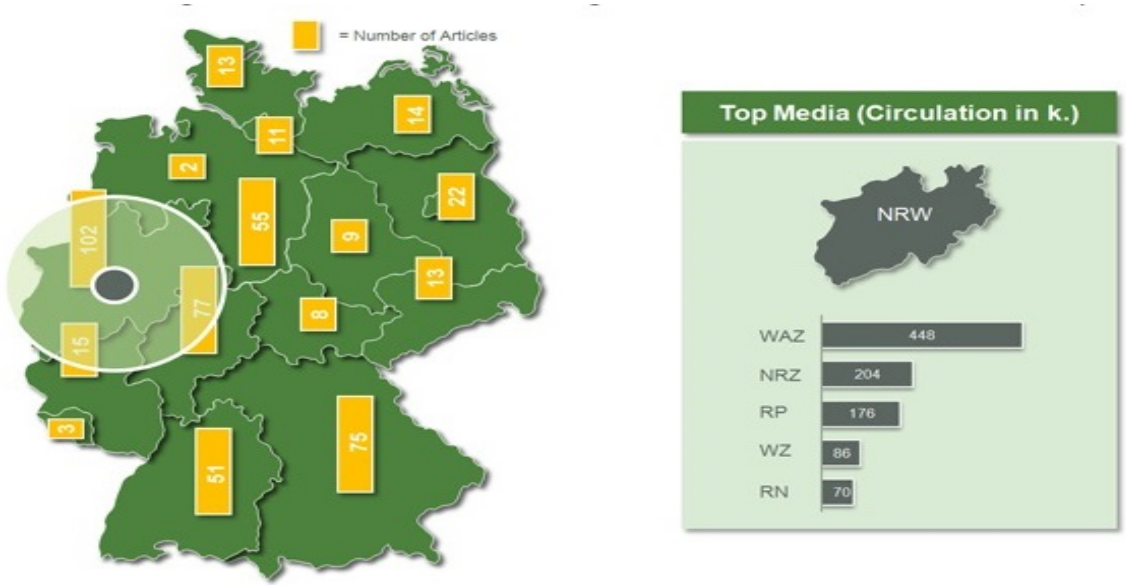


Figure 4: This is the presence analysis

4.1.4 Reputation analysis

A reputation analysis examines the corporate image conveyed to the public by the media. Corporate reputation in the published opinion is measured by using criteria such as economic performance, social responsibility, quality of products and services as well as quality of management and leadership. For mapping the corporate reputation at a certain point in time, pressrelations determines a reputation value, namely the "reputation media score". This calculation method is not only based on the media presence of single reputation criteria, but also includes the reputation relevance of variables as a weighting factor. The reputation media score is described in section 3.1.2.

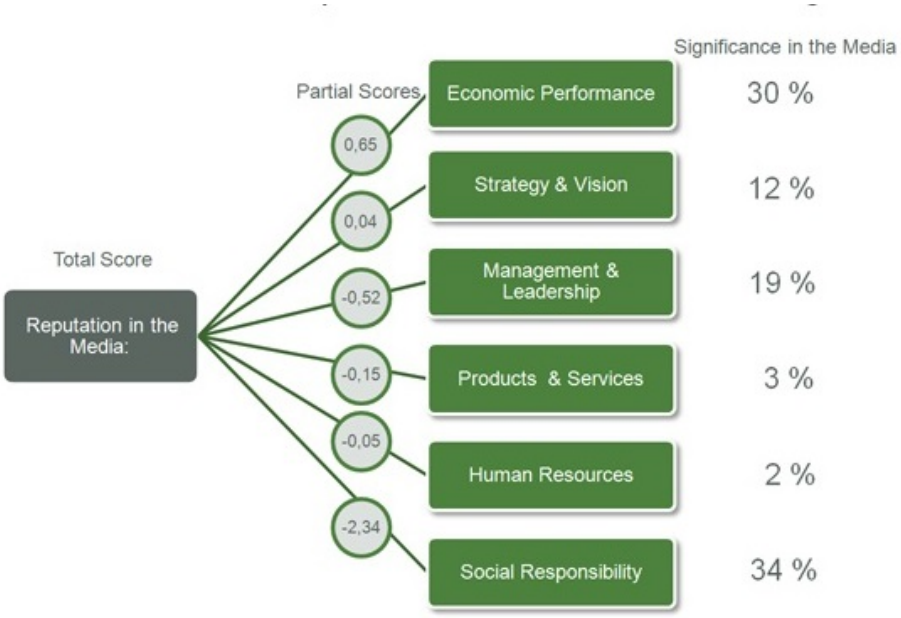


Figure 5: This is the reputation analysis

4.1.5 Issues Analysis

Media coverage not only conveys existing positions and claims, but is also able to reinforce them. An issues analysis reveals to what extent different positions on a certain issue have been picked up by the media. It also determines which information on a certain issue was accessible to which audience. In addition, an issues analysis provides information both on how a company is positioned and how it is perceived. It is mainly carried out to examine critical topics, so-called "hot issues", and can also serve as a reliable early warning system to support the issues management when carried out on an on-going basis.

An issues analysis examines:

- how issues evolve over defined periods of time
- which opinion leaders or stakeholders pay attention to the issue
- which positions and arguments are held by opinion leaders or stakeholders
- to what extent arguments related to other issues or events are linked to an issue
- to what extent a company is associated with a certain issue

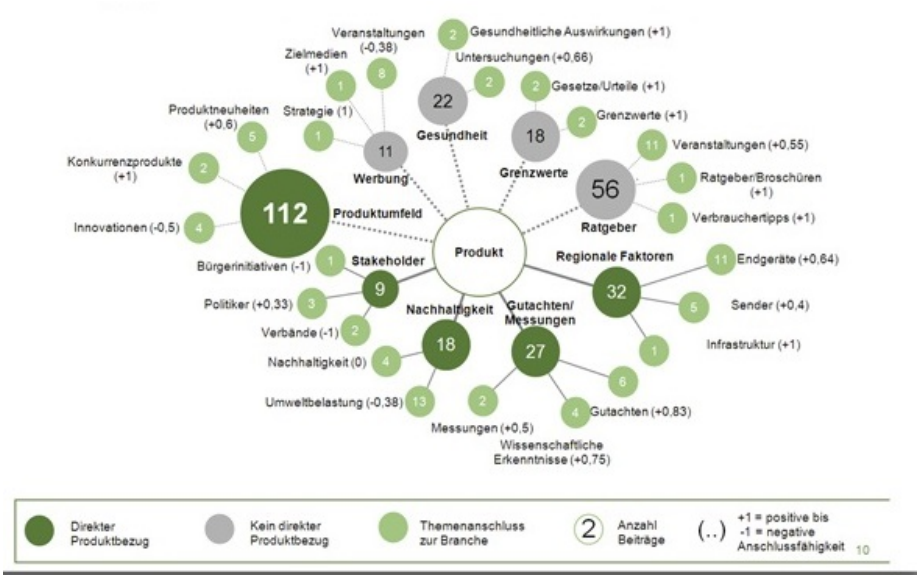


Figure 6: This is the issues analysis

4.1.6 **Stakeholder analysis**

Attention by the media for a certain issue is determined by the number of stakeholders, their networking and campaign abilities and their competence. The stronger an issue is pushed by pressure groups, the longer it will remain on the media agenda. As a result, it is crucial to know all relevant stakeholders and their respective positions.

A stakeholder analysis examines

- which stakeholders take a position on an issue
- which stakeholders emerge as opinion leaders in the media
- in which media arenas these stakeholders are found
- Which positions the stakeholders assume on a particular issue.

The stakeholder analysis provides you with an in-depth overview of all internal and external stakeholders, which determine the media coverage on a certain issue. The statements of each stakeholder group are summarized in the form of central positions and their penetration in the analyses media.

Figure 7: This is the stakeholder analysis

4.1.7 **Advertising Value Equivalency analysis**

In order to ascertain the advertising value equivalency – also known as "editorial value" – of any given news item, we determine the equivalent monetary value of a comparable advertisement, based on the size of the editorial report. The calculation of advertisement rates in print media is based on the cost for a full-page ad. With respect to online media, the "thousand-contact-price" (TCP) for a full banner and the number of visits (unique visitors) serve as essential parameters for calculation. As far as TV and radio are concerned, the length of the news report is used to define its "editorial value".

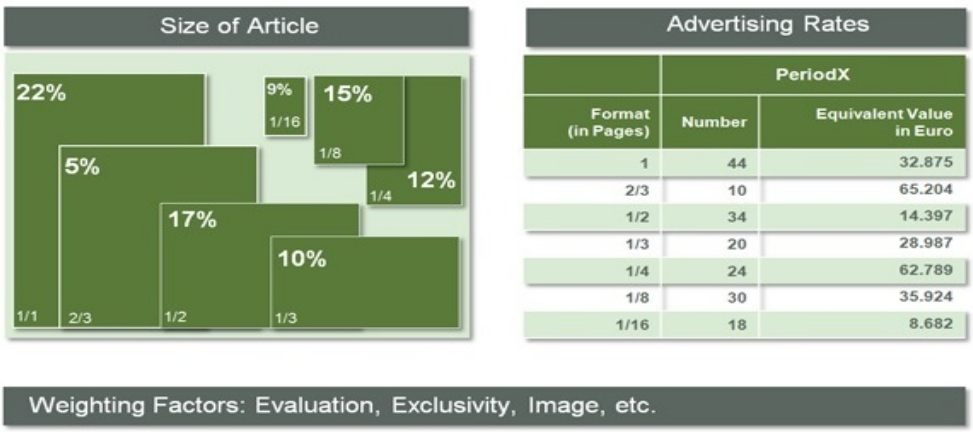


Figure 8: This is the Advertising Value Equivalence analysis

For each of the different monitoring campaigns described above, the set of relevant news items is analysed with respect to different indicators. Not all indicators are relevant for all analyses, and usually the client defines the parameters that have to be taken into account. In the following sections, we will describe the indicators that can be measured and evaluated by media monitoring services.

The following section will describe the indicators that are used in the analysis of traditional media like print and online news, while section 4.3 deals with the set of indicators used in a social media analysis campaign.

4.2 Indicators in Traditional and Social Media

Below we show a summary of indicators in Traditional and Social Media. As it is shown in Figure 9, these indicators are split in: a) Information per medium, b) information per monitoring campaing and c) analysis-central indicators.

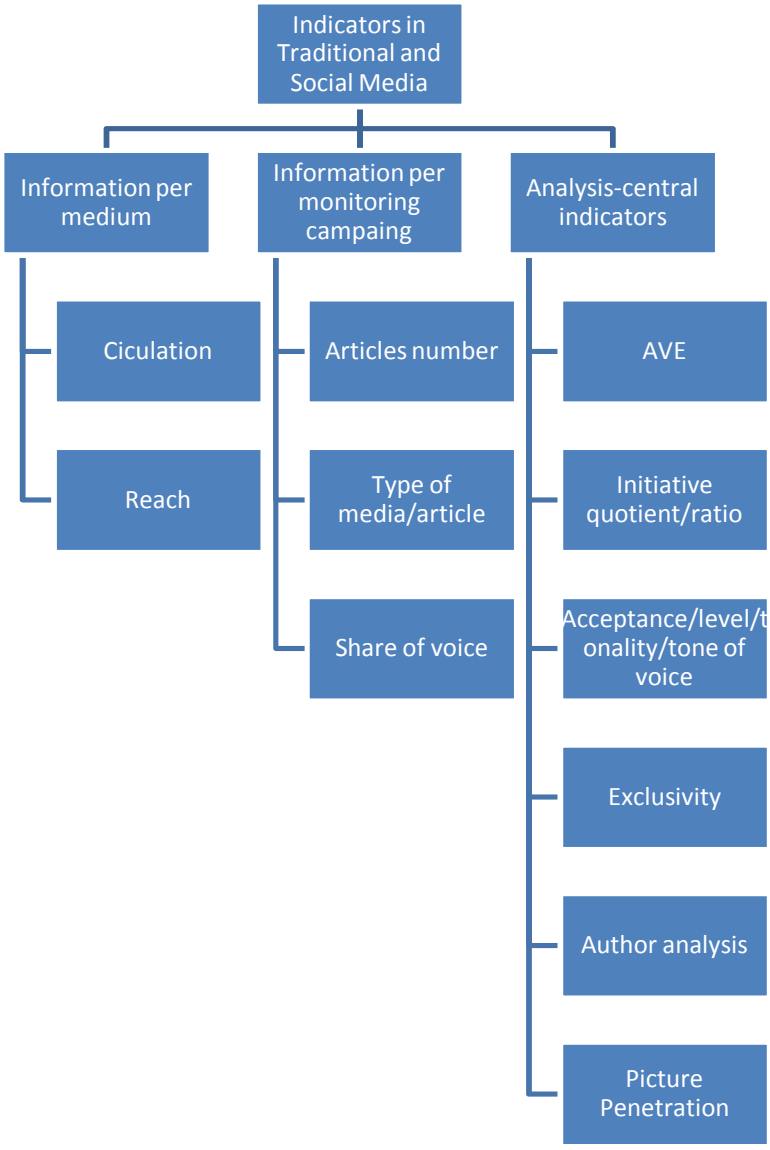


Figure 9: Indicators in traditional and social media

4.2.1 Information per medium

Numbers of copies printed/circulation

For each medium, either a print or an online source, information about the edition is processed. It is usually distinguished between printed and circulated edition numbers. The printed edition specifies the number of units produced in a printing operation. The

circulation includes only the units sold through retail and subscriptions, including distributed free copies.

Information about circulation and copies of printed media is usually distributed by the publishers. For online media, the number of visits to the website per month is included. Per medium, pressrelations stores information about its circulation. For German, the list is exhaustive and the information is delivered attached to the particular news item, for the other languages, the values would have to be researched and incorporated into the system at the point where a particular monitoring campaign is started. Usually, media monitoring companies have media departments, which are responsible for keeping the medium-related pieces of information updated.

Reach

The number of copies does not say anything about the actual number of people who read the particular medium. This information is handled in the indicator 'reach'.

The reach indicates how many readers will actually consume a medium. It is assumed that a copy is usually read by more than one individual; hence the numbers of the circulated editions are multiplied with an average factor for each media type. There are several distributers of reach values per medium. For the German sector, there are two in particular, namely the AllensbachWerbeträgerAnalyse (AWA),[2] and the Arbeitsgemeinschaft Media-Analyse (AGMA).[3] Both agencies investigate the average reach of a large set of different media types and distribute the results. At pressrelations, the factor for each print medium is calculated separately (based on the published values by AWA and AGMA), and per media type the multiplication factor varies.

c_medtype	n_factor_reach
Advertiser	1.80
Specialised press	2.50
Customer magazine	1.50
Consumer press	3.20
Regional daily newspaper	2.50
National daily newspaper	2.50
Magazine	2.90
Inserts	1.20
Sunday newspaper	2.80

Table 5: Reach factor per media type

Hence, for each medium the reach is calculated by multiplying its distribution (made available by the publisher) with the specific factor for the media type.

[2] http://www.ifd-allensbach.de/awa/

[3] https://www.agma-mmc.de/

Information about the reach of radio and TV channels, as well as the reach of one particular broadcast, is distributed by rating agencies that are specialized on audience measurement. For the German sector, e.g., reach and ratings are available from media control[4]. For the UK, the relevant ratings and information about reach is distributed by BARB[5], the Broadcast Audience Research Board. For the French media, the relevant information can be found at Mediametrie[6], Spain is covered by Infortecnica[7] and Bulgarian RTV ratings are distributed by Taylor Nelson Sofres[8].

The reach of an article is calculated on the fly by multiplying the circulation with the appropriate factor for the specific media type. One possible factoring model is:

Circulation × factor (media type) = value

The weighting of the different media types needs to be accessible, since the values can vary, depending on the analyst and the monitoring campaign.

4.2.2 Information per monitoring campaign

Absolute article numbers

Every media monitoring campaign starts with an automatic search for a predefined set of keywords. In an iterative process, the search results are reduced with respect to different criteria, until eventually the final set of news items that feeds into the specific analysis is established. The most basic indicator in a media monitoring campaign covers the accumulated number of news items from all media sources for one specific company, competitor, person, product, etc. Usually, the absolute article number for one particular period is correlated with the number of a comparable period, in order to see whether media interest has increased or decreased during the most recent monitored time frame. Article numbers are the central measure for mostly quantitative analyses like the presence analysis.

Usually, the absolute article numbers are translated into the more revealing indicators like reach. Nevertheless, information about the absolute media coverage is important to almost all clients, particularly in relationship to previous periods.

In figure 10, we see charts from a monthly presence analysis: In print media, 169 articles were published with a total number of 34.82 million copies. For online media, 214 articles were found on websites, which have been visited 686.28 million times. 4 radio and TV broadcasts have been found, with an accumulated reach of 1.45 million. The article number is also an important metric in social media. In this particular campaign, social media items are not further specified for their reach, but it is possible to find a comparable value regarding the reach of e. g. the interactions with these posts are calculated.

[4] http://www.tv-ratings.de/?page=start

[5] http://www.barb.co.uk/

[6] http://www.mediametrie.com/

[7] http://www.infortecnica.com/index.html

[8] http://www.tnsglobal.com/

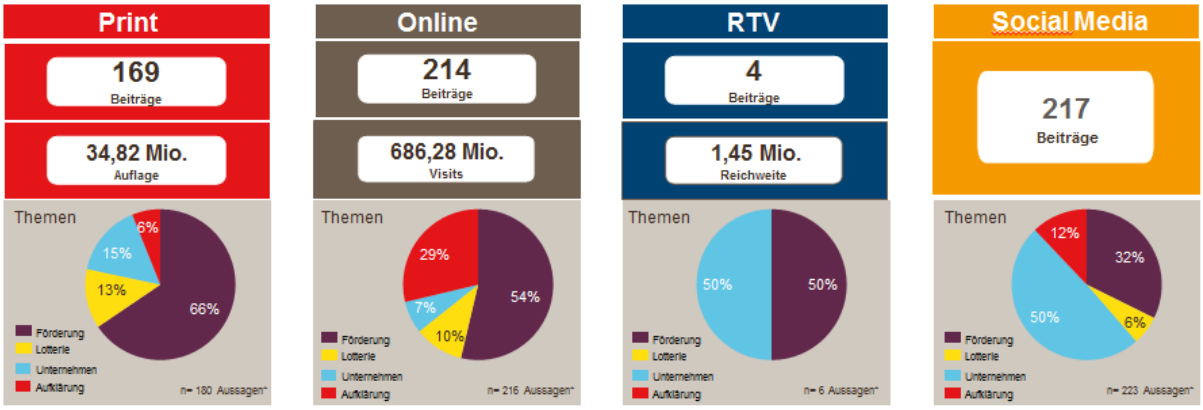


Figure 10: This is the Absolute numbers, circulation, and reach combined

Type of media, type of article

The absolute number of news items is usually subdivided into different types of media on the one hand, and different types of articles on the other, since not all types of contributions are equally important for a client. The "type of article" indicator determines if the coverage comes in the form of an interview, comment, feature, press release etc. The "types of media" indicator covers categories like daily newspaper, magazine, specialized press, etc.

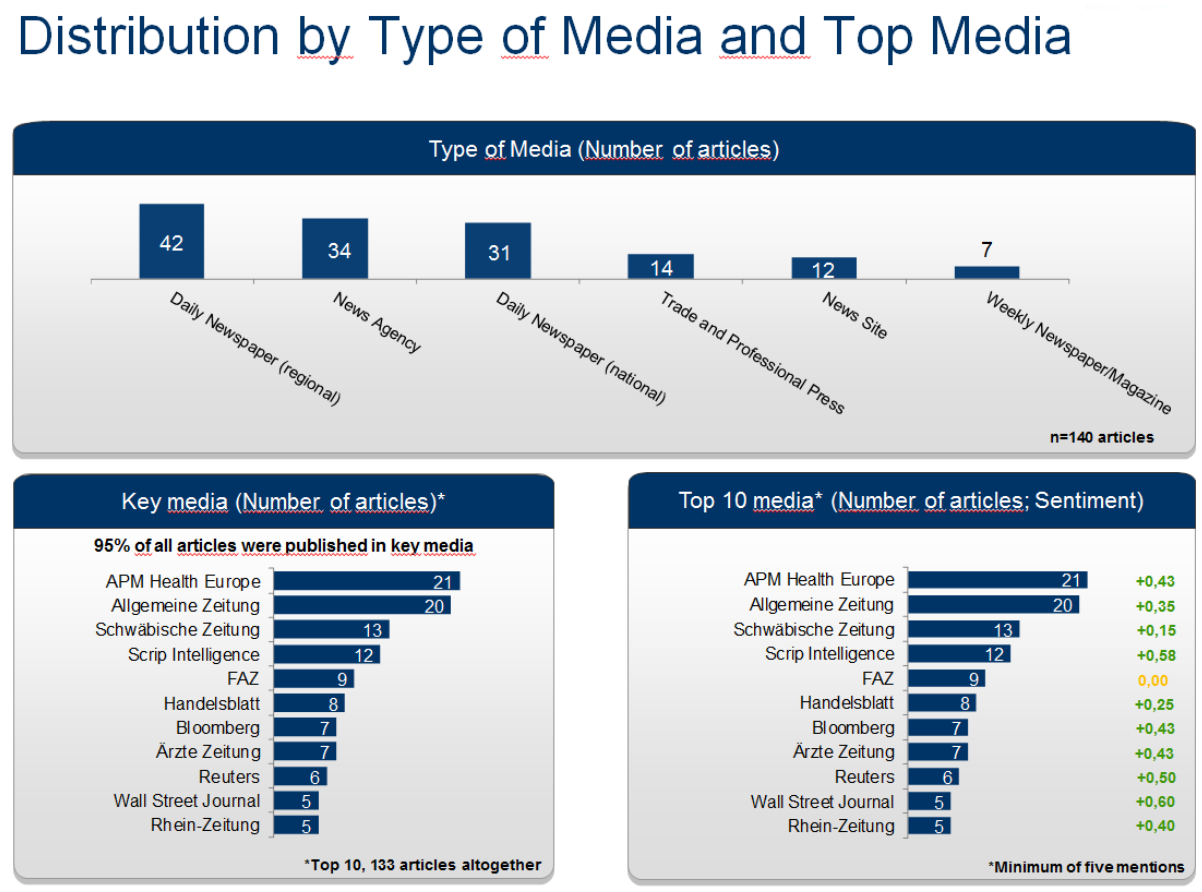


Figure 11: Classification by type of media

Share of voice

In traditional as well as social media, the share of voice indicates the percentage of news items about a particular focus topic (e.g., company or product) in the overall media coverage on the given topic. For example, if the overall amount of articles on the topic "household appliances" is 100, with 25 articles being about a specific manufacturer Y, then the share of voice of manufacturer Y in the discussion of household appliances is 25 percent.

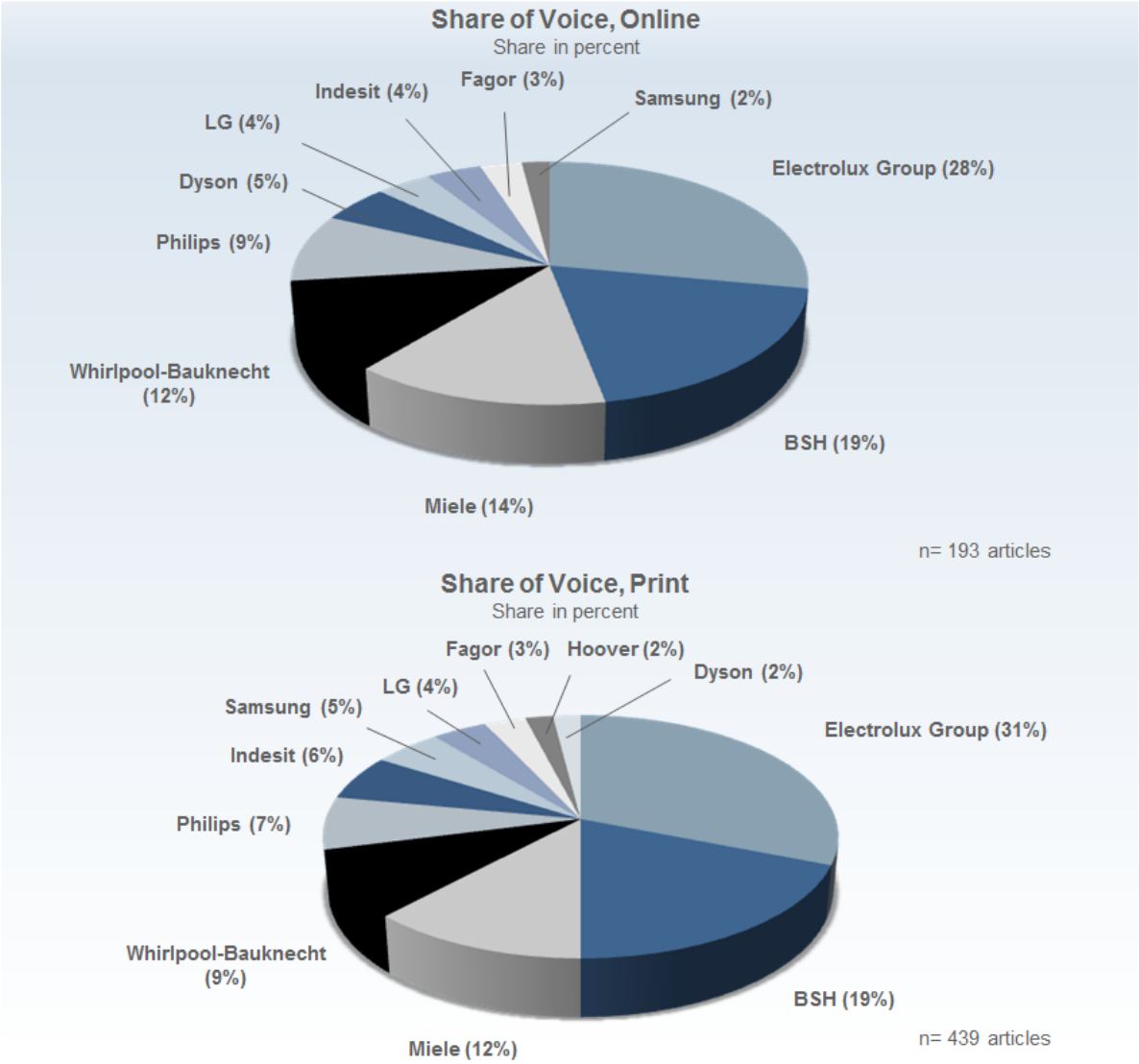


Figure 12: This is the share of voice for household appliances manufacturers (from NewsRadar)

Pressrelations calculates the share of voice by dividing the object mentions by the total mentions.

 # Object mentions / (total mentions (object + competitor A, B, C…n))

4.2.3 Analysis-central indicators

AVE (Advertising Value Equivalency)

The AVE is a highly debated indicator in PR and media monitoring. Although the rejection of the AVE is a central part of the Barcelona principles,[9] a declaration of media measurement principles that has been published in 2010 by AMEC, the international association for the measurement and evaluation of communication, the calculation of the AVE is still demanded by the majority of media monitoring clients. Roughly speaking, the AVE calculates the cost of an advertisement the size (for print) or length (for RTV) of the measured news item. We distinguish between weighted and unweighted AVEs.

The unweighted AVE of print articles is determined on the basis of the article size/thread length and the display price for the relevant medium. The unweighted AVE of online articles is calculated on the factors visits (users per month) and the full banner price for thousand contacts. It indicates the costs that would arise for the implementation of an advertising campaign reaching 1000 users in this medium. The unweighted AVE for RTV contributions is calculated based on the ad price per second for each individual transmission date (DD/HH/MM).

For the calculation of weighted AVEs, a multiplication factor is determined. This is done in order to account for the fact that certain publications are more credible than others, just as certain publication are more central to a particular client than others. To give one example, an article about a new car will have more impact when it is printed in a specific magazines about cars than it would be in a random local newspaper, since it can be assumed that the readers of the car magazine have a more concrete interest in cars than the majority of the readers of a local newspaper. Hence, an article about the new car has a bigger advertising equivalency, which can be factored in by multiplying the AVE of the particular magazine with a predetermined factor.

Another reason for assigning a weight to the AVE can be the idea that generally, a news message can be seen as more valuable than an advertisement, since it is considered more independent and objective than an ad that has been placed by the company itself. This type of factor is called the “PR factor” (Jeffries-Fox, 2003).

Just like the circulation values, the unweighted AVE is part of the Meta data for each print or online news item that is fed into the MULTISENSOR repository by pressrelations. In order to allow for a weighted AVE, there needs to be a possibility for the analyst to assign weights when a campaign is set up.

[9] http://amecorg.com/2012/06/barcelona-declaration-of-measurement-principles/

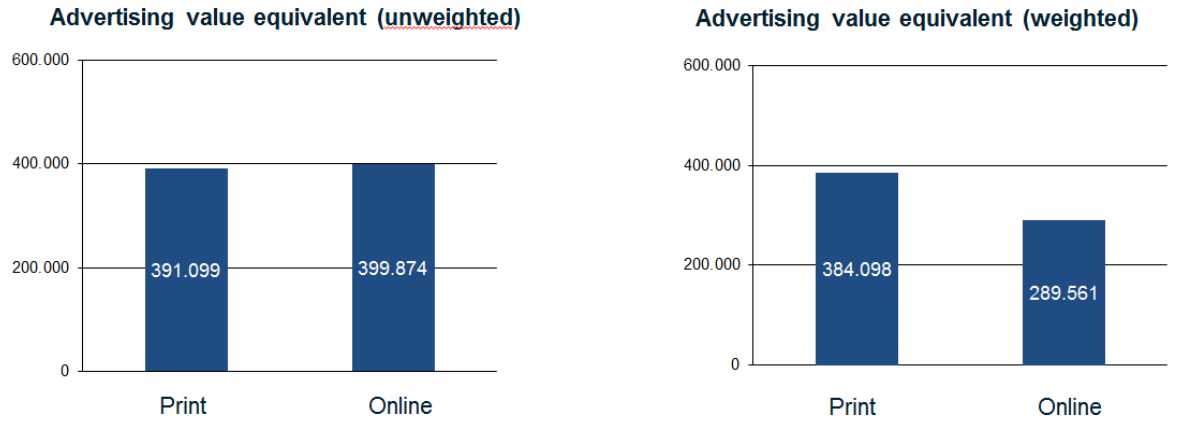


Figure 13: This is the weighted and unweighted AVE for the same set of news

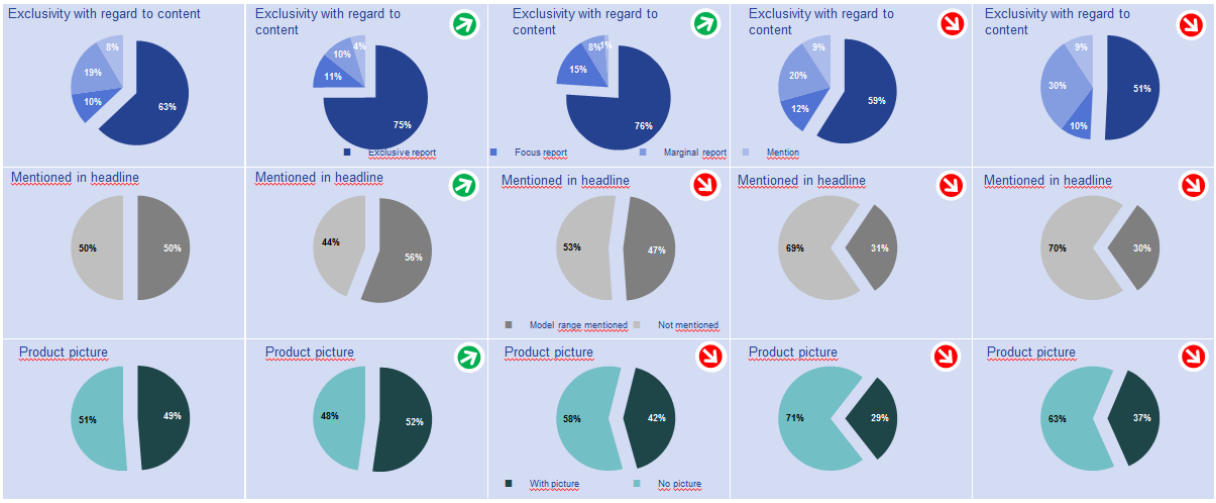


Figure 14: This is the AVEs for 5 different products for the same set of news

Initiative quotient/ratio

The initiative quotient or ratio indicates the proportion of the articles (in percent) that contain a recognizable reference to a press activity of the company (press release, press conference, other PR events). A high quotient initiative speaks for successful public relations of the company. This quotient is of course applicable for traditional and social media.

 # PR-initiated mentions / (total mentions (self-initiated + externally initiated mentions))

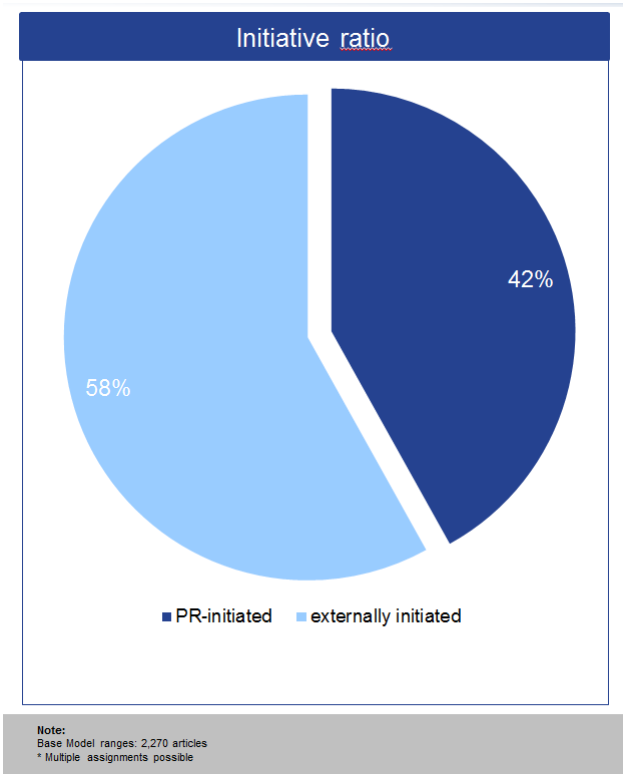


Figure 15: This is the initiative ratio

Acceptance level/Tonality/Tone of Voice

The sentiment of an expression, a sentence or an article can vary from positive to neutral to negative. Usually, media monitoring clients not only want to know how much media coverage has been found for a particular period, they also want to know whether what has been said or written is positive or negative. Hence, in the analysis, statements are assessed from the perspective of the analysis object, e.g. a company, person or product. The statement can be implicit or explicit and the tonality is usually measured on a numeric scale. Tonality can either be measured on sentence or statement level, or, more globally, on article level. In social media the share of "neutral" posts is often lower than in traditional media because usually people want to express their opinion and their approval or rejection regarding a certain product, company or topic while traditional media often provides neutral information.

Pressrelations measures the acceptance level by dividing the sum of the tonality values by the number of measured items.

$$\sum \textit{tonality values} / \textit{\# measured items}$$

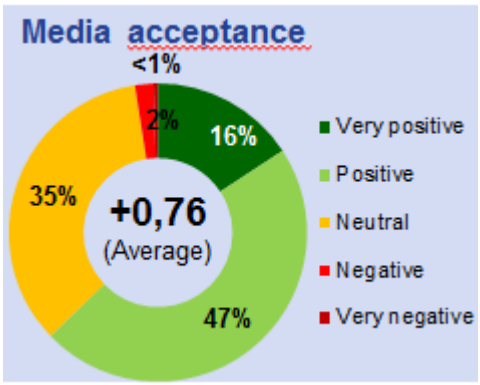


Figure 16: This is the acceptance level

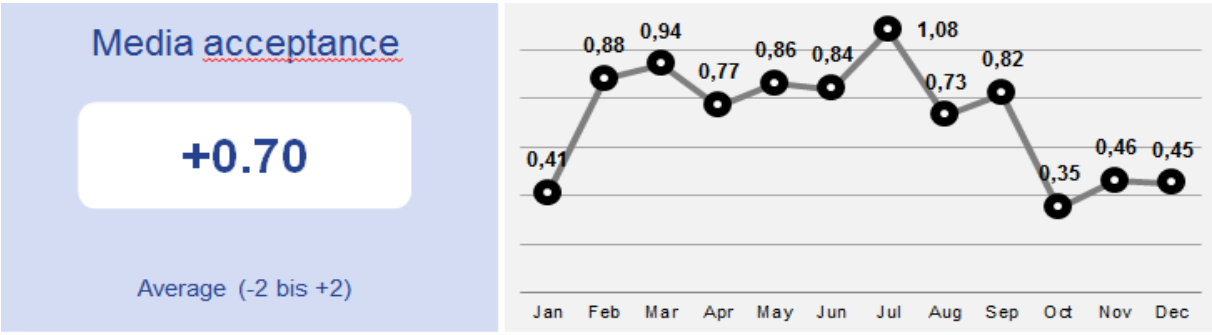


Figure 17: This is the acceptance level in a temporal dimension

Exclusivity

This indicator refers to the centrality of a news item from the point of view of the analysis object. For every article, it is determined whether it is an exclusive report, a focus report, a marginal report or a marginal mention of the analysis object. This indicator can be combined with the initiative quotient in order to evaluate the impact of a client's PR campaign (Figure 18) shows the overall exclusivity measure of a specific campaign, subdivided into two different types of contents: company-initiated and media-initiated content (initiative quotient). Exclusivity is more of a traditional media indicator because social media posts are often quite specific and short (with the exception of blogs) and therefore nearly always "exclusive" regarding a certain topic.

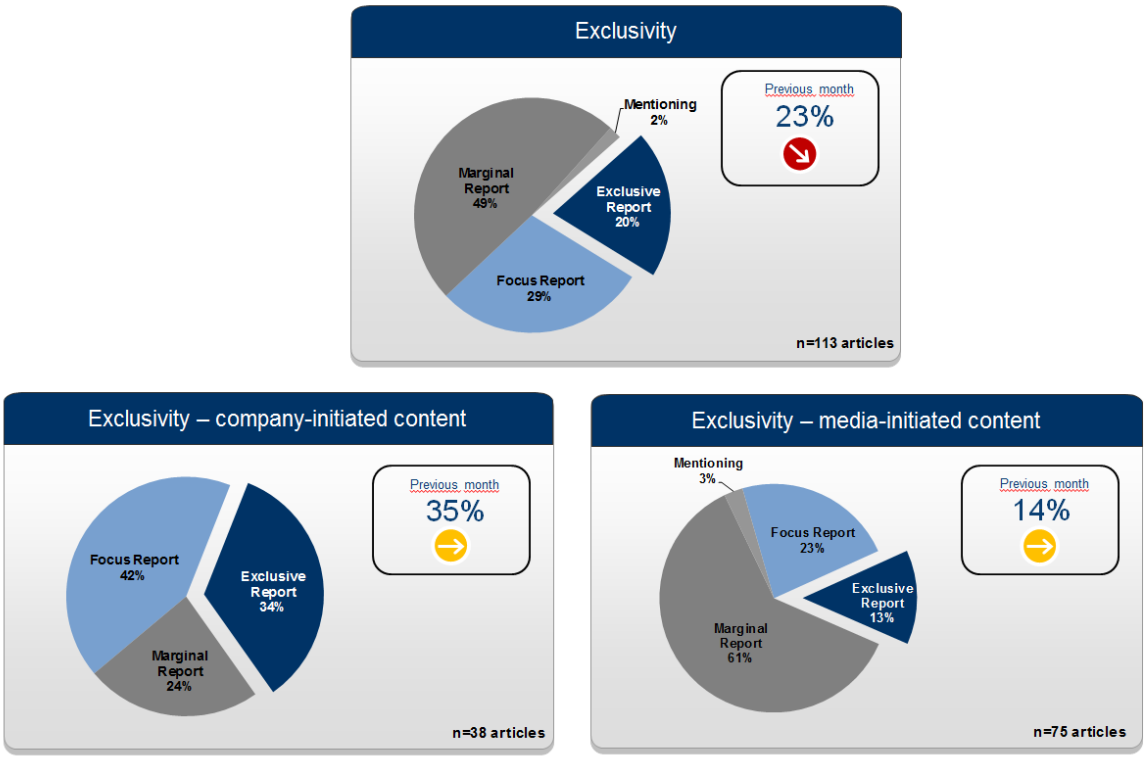
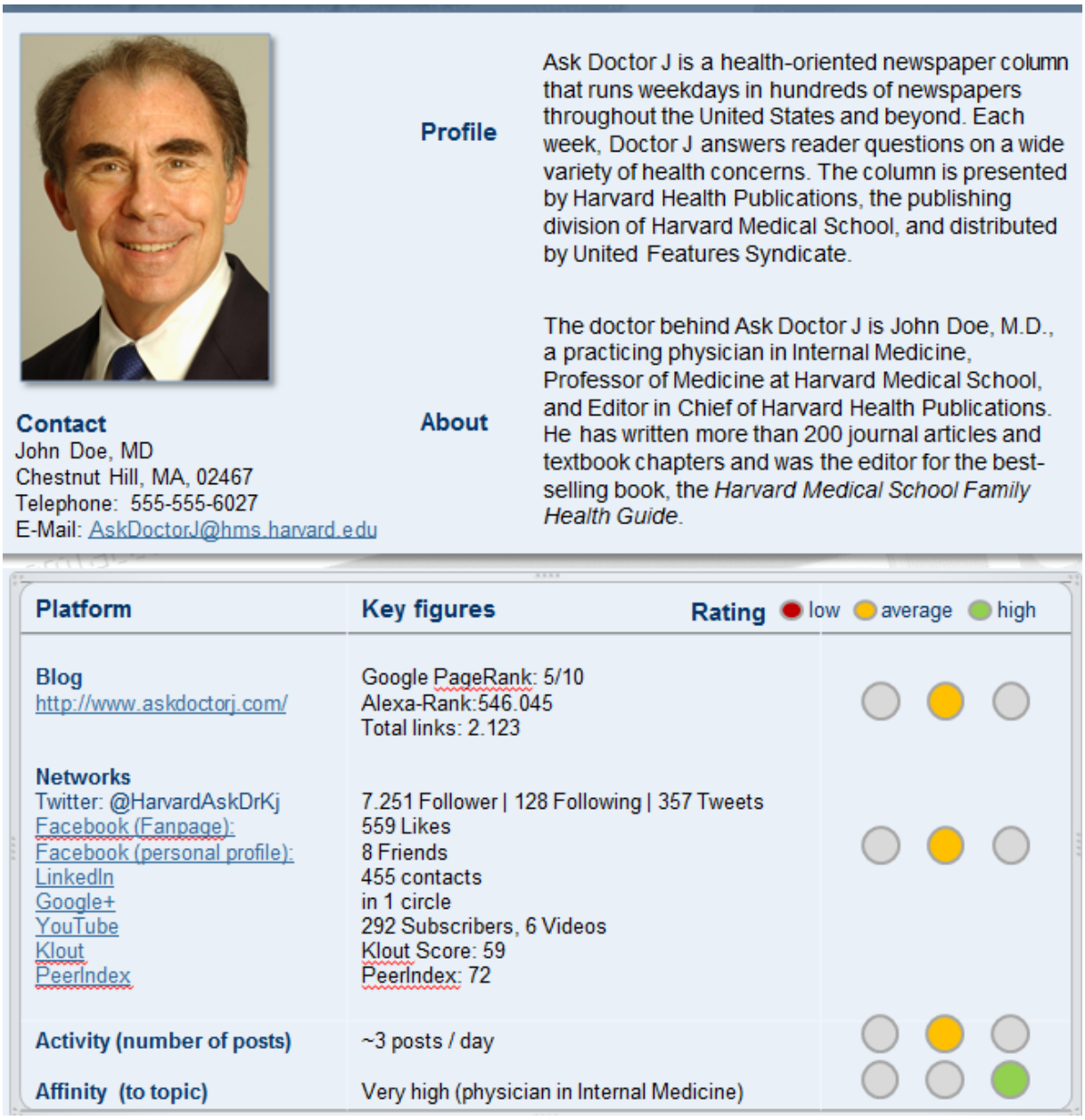


Figure 18: This is the exclusivity and initiative ratio

Author analysis

Many media monitoring clients are interested in detailed information about their stakeholders and opinion leaders in their specific field, be it because it is well-known that a particular author is known as a key influencer and his or her contributions have high impact, or because they would like to get in touch with single authors. For that reason, many media resonance analyses offer additional information about key influencers or outstanding journalists. The author information contains publicly available data like social media identities; publication channels the author uses, contact details (if available), etc.



Profile

Ask Doctor J is a health-oriented newspaper column that runs weekdays in hundreds of newspapers throughout the United States and beyond. Each week, Doctor J answers reader questions on a wide variety of health concerns. The column is presented by Harvard Health Publications, the publishing division of Harvard Medical School, and distributed by United Features Syndicate.

About

The doctor behind Ask Doctor J is John Doe, M.D., a practicing physician in Internal Medicine, Professor of Medicine at Harvard Medical School, and Editor in Chief of Harvard Health Publications. He has written more than 200 journal articles and textbook chapters and was the editor for the best-selling book, the *Harvard Medical School Family Health Guide*.

Contact
John Doe, MD
Chestnut Hill, MA, 02467
Telephone: 555-555-6027
E-Mail: AskDoctorJ@hms.harvard.edu

Platform	Key figures	Rating: low / average / high
Blog http://www.askdoctorj.com/	Google PageRank: 5/10 Alexa-Rank:546.045 Total links: 2.123	
Networks		
Twitter: @HarvardAskDrKj	7.251 Follower | 128 Following | 357 Tweets	
Facebook (Fanpage):	559 Likes	
Facebook (personal profile):	8 Friends	
LinkedIn	455 contacts	
Google+	in 1 circle	
YouTube	292 Subscribers, 6 Videos	
Klout	Klout Score: 59	
PeerIndex	PeerIndex: 72	
Activity (number of posts)	~3 posts / day	
Affinity (to topic)	Very high (physician in Internal Medicine)	

Figure 19: This is the author information

Picture Penetration

Media monitoring campaigns, which focus on particular brands or products, are usually interested in media coverage that resembles some form of advertising value. Particularly for sponsorships, the frequency with which a logo or the name of a product or brand is visible; is the key indicator for this campaign. One way to gain insights is to measure the positive exposure of products, logos, etc. in pictures or TV broadcasts. In pressrelations analyses, it is distinguished between “logo completely visible” and “logo partially visible”, but there are also other approaches e.g. working with percentage quotation. Although this indicator is also measurable in social media it is predominantly used in traditional media analyses.

At pressrelations, there are two different ways to determine the picture penetration of logos, etc. Usually, the analysis is done manually, but in particular cases, this information can be bought from the supplier of the medium.

4.3 Base of calculation in Social Media

This section will focus on potentially relevant criteria for the assessment of the importance of individual social media posts. These will be determined per channel and we will then try to find an indicator that ensures comparability. As there currently is no industry standard, we will have a look at different models that have been proposed. Network analysis and detection of influencers will play a part in this.

There are 3 levels of indicators in social media: a) metrics, b) key performance indicators and c) scores. The abstraction of data increases with each level:

4.3.1 Metrics

The metrics are values that can be directly measured within a certain social media channel. Each channel has its own metrics and the majority of them are quantitative. The metrics are pure values without any weighting or correlation. Examples are the number of Followers of a Twitter user, the number of tweets but also values like the user's gender.

In the following examples are provided separated for each social network.

Twitter

Follower: The number of Twitter users who subscribed to a user's Twitter channel to see his posts in the Twitter news stream.

Following: The number of Twitter users a user has subscribed to.

Number of Tweets: The number of Tweets a user has published on Twitter.

Location (User): This is a free text field, which can be used to show the usual habitation of the user.

Location (Tweet): Shows (if activated) where a certain Tweet has been published.

Gender: This field shows the gender of the user.

Language: Optional information of the user regarding his preferred language.

Retweet: Users on Twitter can repost a certain Tweet in their own timeline by clicking on the "Retweet" button so their followers are able to see this tweet in their news streams.

Favourite: If a user doesn't want to re-post or to comment on a certain Tweet but still wants to express that he likes, approves or agrees with the content, he can click on the "Favourite" button under each Tweet.

Facebook

Fans: The number of fans a certain Facebook fanpage has. If a user becomes a "Fan" of a certain fanpage he shows that he likes it. Usually he also agrees to subscribe to this fanpage to see new posts in his Facebook news stream but the subscription is optional.

Friends: The number of people a user has added to his network on Facebook. Users on Facebook can restrict their posts to all friends or parts of their network so these posts can only be read by the chosen group of people. It is also possible to decide whether personal user information should be visible to everyone or just your friend network on Facebook.

Likes: Users can express that they like, approving or agreeing with a fanpage, a post or a comment by clicking on the "Like" button.

Shares: If users click on the "share" button of a certain post they repost it on their own timeline so their friends and followers can see this content. A user can share a post as it is or modify the content before reposting it.

Comments: Users can comment on most of the content on Facebook and therefore express their opinion and discuss with other users about a certain post. It is dependant of the user's or the fanpage's options whether it is necessary to be a friend or fan to be able to comment.

Comment-Likes: Users can also express that they like a certain comment by clicking on the "like"-Button.

Gender: Facebook allows users to specify their gender. It is not only possible to choose between male and female but also to create your own gender description.

Location: It is possible to specify your habitation on Facebook and to decide whether this information (as all the other personal information) should be visible to everyone or just to your friends on the network.

Language: Users can specify here which languages they speak.

4.3.2 Key Performance Indicators

The key performance indicators (KPIs) are the result of two or more metrics of certain channel that are set in a relation respectively correlation to each other. This is necessary to compare the relevance/performance of users/channels within a single social network. KPIs help to keep a better overview over a channel by simplifying the available metrics to a few correlations. Due to the fact that KPIs are always based on a generalisation of the relation between the channel's metrics it is possible that these KPIs are inaccurate for single users/channels with special conditions.

Examples per social network are provided below.

Twitter

Follower-Following-Ratio

 # Followers / # Following = [value]
The number of followers is divided by the number of users this account is following.
A value >1 means that the user seems to publish interesting information
A value <1 means that the user is more of a "listener"

Twitter interaction rate per Tweet

 # Retweets of a Tweet + # Favourites of a Tweet / # Followers of the user = [value]
The number of interactions with the Tweet (Retweets and) are put in proportion to the number of followers of the author who are potential readers of his content. The higher the value the higher is the interaction rate with the Tweet.

Facebook

Facebook Fanpages - interaction rate per post

 # Likes + # Comments + # Shares / # Followers = [value]
The number of interactions with the post (Likes, Comments, Shares) are put in proportion to the number of Fans of the fanpage who are the potential readers of the content. The higher the value the higher is the interaction with the post. The value needs to be compared either with itself over time or with other fanpages to see how good or bad the interaction rate is developing.

User activity

 # Posts / {period} = [value]
The number of posts a user has published are put into proportion of a certain period, e. g. a day, a week, a month etc. The value needs to be compared either with itself over time or with other users to see how active a user is or how his activity level is changing over time.

4.3.3 Scores

Scores are the highest level of data abstraction. The purpose of a score is to normalise the different metrics and KPIs of each channel to evolve comparable values. Therefore the comparable values of each channel are converted into a scoring system with equal scales. Even though the channels have many comparable values it is possible that their relevance to the channel differs because the usage patterns within the single social media channels can be very different from each other. This fact diminishes the reliability of scores to some point but not that much that they are useless. They help to simplify the relevance evaluation between the many different social media channels and to get a big overview. The values are not necessarily compared directly, it is possible that based on experience or individual relevance biases, and the results of each channel are increased or decreased by a certain factor.

An Example for a Score is “interaction” where the e. g. the number of Retweets, Replies and Favourites of a certain Tweet are set in relation to the Followers of the user and this value is compared with e. g. the result of Likes, Shares and Comments of a certain Facebook status which is set into relation to the Friends/Fans.

Total Interaction Rate:

[TWITTER]((Favourites + Retweets) / Follower) + [FACEBOOK]((Likes + Comments + Shares) / Likes|Friends) + [YOUTUBE](Comments + Likes + Dislikes) / Views) + [BLOGS]((Comments + Social Shares) / Backlinks) + … = {value}

The total interaction rate combines the proportion of interactions with the number of potential readers for each channel to one single interaction rate. The principle is the same for each channel – every measurable act of interaction with a post is counted, maybe amplified or mitigated due to its importance and informative value and then put into proportion to the current audience. Audience means the measurable number of people who are likely to have seen the post. The scale of the interaction rate goes from 0 to infinite. The higher the score the higher is the weighted interaction with the content.

5 INDICATORS FOR SME INTERNATIONALISATION

In most of the cases the problem of internationalisation for SMEs is a problem of comparing different countries that could be potentially interesting for exporting products. The selection of the correct country for the international investment depends on a number of indicators. These indicators are very important in order to make a first analysis of the different options that can be presented in a global market. Once the global market is studied we can narrow down in terms of number of the selected countries. To do that, first of all we have to conduct an initial screening based on the different criteria, as geographical proximity, financial situation of the country and political stability, growth of the country's GDP, market size, barriers, etc. All these indicators have to be related with the portfolio of the products of the company.

In order to analyse and organize all the indicators needed for the SME, the best way is to categorize by considering a PESTL (Political, Economic, Social, Technological, Legal) Analysis. The reason to select PESTL analysis is that the latter describes a framework of macro-environmental criteria, which is currently considered in the strategic management of SMEs in order to gather of opportunities or threats. In this case we extended the more broadly accepted PEST analysis by adding the legal factor (L), because is very important for the companies to be aware about laws and the updates of it, given the fact that the legislation is an important factor to decide or not to export in a specific country.

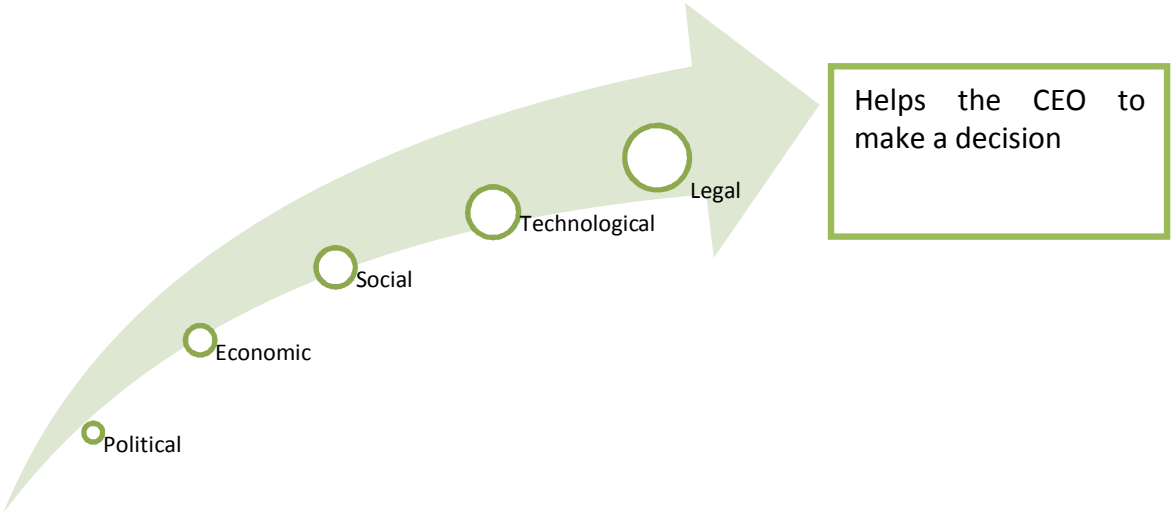


Figure 20: PESTL analysis

5.1 PESTL Analysis

In order to cover all the indicators for internationalisation, a SME studies the factors so as to understand its potential consumers and market requiremetns.

The following overview shows the indicators for the SME's organised by PESTL factors.

5.1.1 Political

Government type

This indicator refers to the government structure. This is important to know how is the government is structured and how does it work.

Government Type	Federal Republic

Chief Executive	Joachim Gauck	Second Chief Executive	Angela Merkel
Type	Fed. President	Type	Chancellor

Branch	Last Election	Term Length
Executive	2010	5 years
Legislative	2009	5 years

Legislature Size	667 members

Figure 21: Government Type

Political instability

The Political instability Index shows the level of threat posed to governments by social protest. This information is important in order to know whether the targeted country is stable in order to invest.

Key: ▲ Risk has increased since 2007 ▼ Risk has decreased since 2007 ▶ Risk is unchanged since 2007
Very high risk | High risk | Moderate risk | Low risk

Rank	Country	Underlying vulnerability	Economic distress	Index score		2007 score
165	Norway	0.4	2.0	1.2	▲	0.2
164	Denmark	0.4	4.0	2.2	▲	0.2
163	Canada	1.7	4.0	2.8	▲	1.8
161	Finland	0.4	6.0	3.2	▲	1.2
161	Sweden	0.4	6.0	3.2	▲	1.2
160	Switzerland	0.8	6.0	3.4	▲	0.4
158	Costa Rica	2.1	5.0	3.5	▲	1.5
158	Mauritius	2.1	5.0	3.5	▲	2.5
154	Australia	1.3	6.0	3.6	▲	0.6
154	Austria	1.3	6.0	3.6	▲	0.6
154	New Zealand	1.3	6.0	3.6	▲	0.6
154	Luxembourg	1.3	6.0	3.6	▲	0.6
153	Czech Republic	3.3	4.0	3.7	▲	2.7
150	Germany	1.7	6.0	3.8	▲	1.8
150	Japan	1.7	6.0	3.8	▲	0.8
150	Slovenia	1.7	6.0	3.8	▲	1.8
149	Oman	3.8	4.0	3.9	▲	2.9
146	Hong Kong	2.1	6.0	4.0	▲	1.0
146	Belgium	2.1	6.0	4.0	▲	2.0

Figure 22: Index of political instability

Corruption perception Index level[10]

The Corruption Perception Index (CPI) ranks countries or territories based on how corrupted a country's public sector is perceived to be. The CPI is important for the SME's to know this Index in order to avoid risks and increase unexpected expenses.

European Union and Western Europe

RANK	COUNTRY	2014 SCORE	2013 SCORE	2012 SCORE
1	Denmark	92	91	90
3	Finland	89	89	90
4	Sweden	87	89	88
5	Norway	86	86	85
5	Switzerland	86	85	86
8	Netherlands	83	83	84
9	Luxembourg	82	80	80
12	Germany	79	78	79
12	Iceland	79	78	82
14	United Kingdom	78	76	74
15	Belgium	76	75	75
17	Ireland	74	72	69
23	Austria	72	69	69
26	Estonia	69	68	64
26	France	69	71	71
31	Cyprus	63	63	66
31	Portugal	63	62	63

Figure 23: This is a Corruption perception Index level

Figure 24: Corruption perception Index

[10] http://country-facts.findthedata.com/l/29/Germany

5.1.2 Economic

The economic indicators are considered as the most important indicators. In the following we go through them in detail.

GDP Country:

This indicator reflects economic power of a country and its evolution over time (evolution of the economy). Countries with a high GDP and increase of their annual economies show strong and steady development.

Figure 25: GDP Country

GDP purchasing power parity (PPP):

PPP shows the purchasing power of the population. This indicator is important to know whether a product can be consumed by the majority of the population or not (improve). A high-PPP displays a high capacity of consumers to buy products at the same time.

Country name	2010	2011	2012	2013
Finland	38,300.8	40,280.3	39,730.4	39,812.4
France	35,873.0	37,312.4	37,114.9	37,871.9
French Polynesia				
Gabon	16,387.6	17,488.5	18,356.3	19,264.3
Gambia, The	1,633.4	1,545.0	1,612.7	1,661.4
Georgia	5,818.2	6,322.5	6,823.0	7,176.4
Germany	39,557.6	42,381.0	43,170.7	44,469.4
Ghana	3,003.0	3,445.7	3,733.8	3,992.1
Greece	28,904.1	26,944.5	25,588.3	25,705.0
Greenland				
Grenada	10,952.0	11,220.9	11,247.1	11,645.3

Figure 26: This is the GDP Purchasing power parity

Customs and tariffs:

This indicator shows the barriers to entry in a particular country for your business development. Very high barriers to entry may dismiss the viability of the project in the country e.g. tax, vat, fees, duties.

EU Tariffs

Search results

Exporting Buttermilk, curdled milk and cream, yogurt, kephir and other fermented or acidified milk and cream, whether or not concentrated or containing added sugar or other sweetening matter or flavoured or containing added fruit, nuts or cocoa - Not flavoured nor containing added fruit, nuts or cocoa from Argentina

Code	Product description
04	DAIRY PRODUCE; BIRDS' EGGS; NATURAL HONEY; EDIBLE PRODUCTS OF ANIMAL ORIGIN, NOT ELSEWHERE SPECIFIED OR INCLUDED
0403	Buttermilk, curdled milk and cream, yogurt, kephir and other fermented or acidified milk and cream, whether or not concentrated or containing added sugar or other sweetening matter or flavoured or containing added fruit, nuts or cocoa
0403 10	-Yogurt
0403 90	-Other
	--Not flavoured nor containing added fruit, nuts or cocoa
0403 90 11	---In powder, granules or other solid forms
0403 90 11	----Not containing added sugar or other sweetening matter, of a fat content, by weight
0403 90 11	-----Not exceeding 1,5 %

Origin	Measure type	Tariff	Conditions	Footnote	EU Law
ERGA OMNES	Third country duty	100.4 EUR/100 kg			R2204/99

Figure 27: Customs and tariffs

Exchange rate

SMEs need to take into account aspects such as the parity euro-dollar, the deflation. We should be aware that an increase in currency prices can increase the price of products and commercial activities and make them no longer economically profitable.

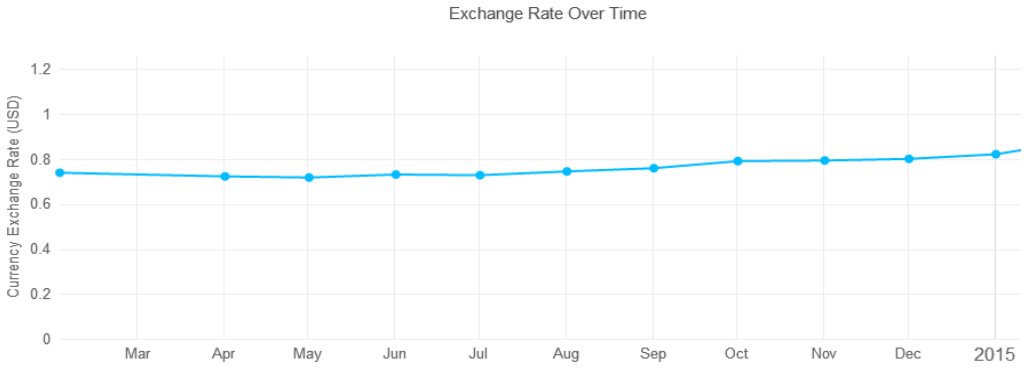


Figure 28: Exchange rate

Consumer Confidence

This indicator measures the degree of confidence that a consumer feels about the state of the economy and their personal financial situation.

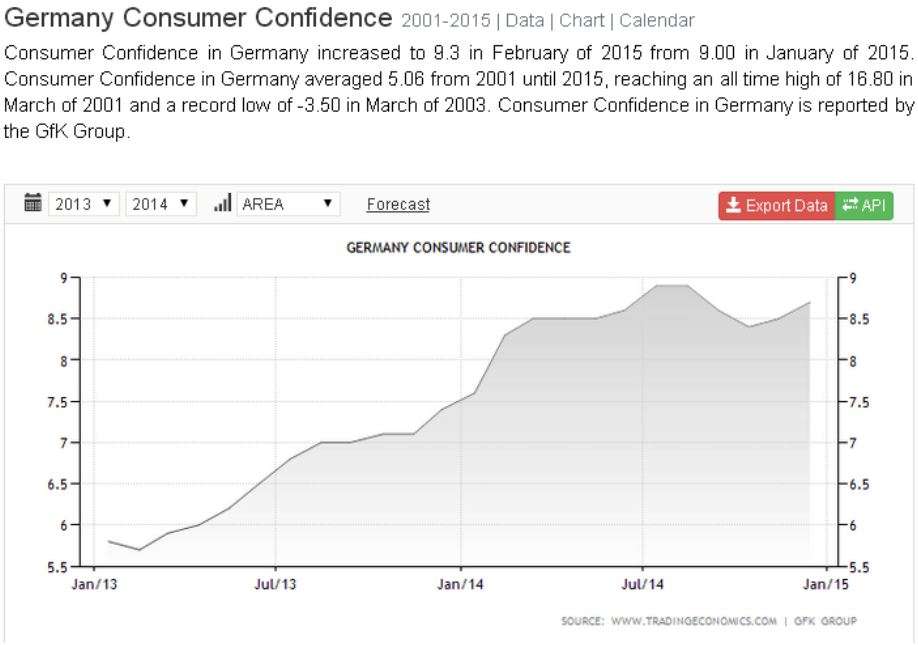


Figure 29: This is the consumer confidence

Business Confidence[11]

This indicator measures the strength of businesses in a specific country.

Figure 30: Business Confidence

[11] http://www.tradingeconomics.com/germany/business-confidence

Export & Import

If a country has high import volumes, it is easier to make business with them because they are familiar with it. On the one hand this could mean that the country doesn't provide similar products but on the other hand there might be many competitors that export.

Figure 31: Example of an export & import indicator

Market share

Market share is the percentage that a company or business owns in a particular market. This indicator is closely related to the market potential and can be calculated as follows:

Unit Market Share (%)=Unit Sales (#)/Total Market Unit Sales (#)

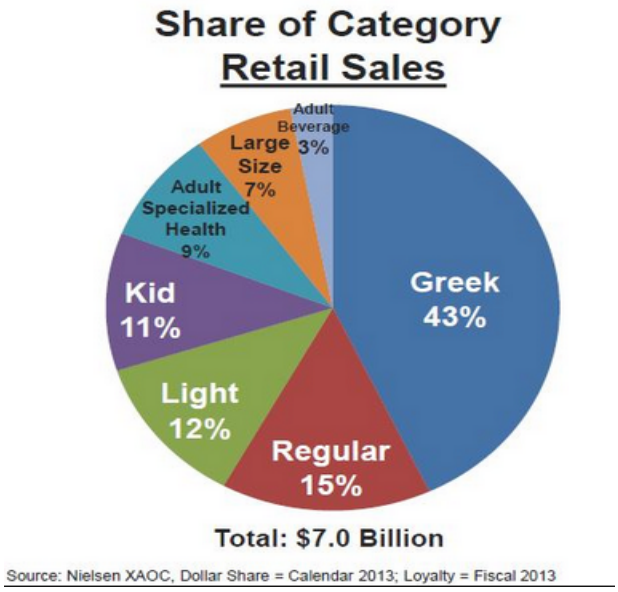


Figure 32: Market share

Picture Penetration and competitors

Usually the company knows already if there are big companies that export similar products internationally. If not, SMEs need to discover the competitors by searching online.

Figure 33: Picture penetration and competitors

Distance between country

The distance between the countries of interest is very important in order to calculate the investment and the days of transportation.

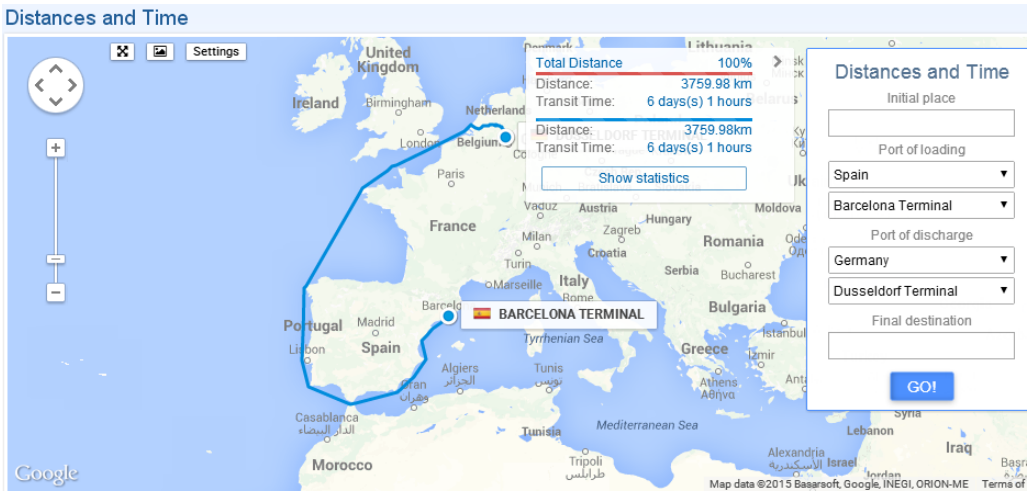


Figure 34: This is the distance between Countries

5.1.3 Social

Social factors relate to the habits of the consumer, as well as the status of the general public.

Human Development Index (HDI)

HDI is used to compare countries development based on 3 dimensions: life expectancy, years of schooling and gross national income per capital.

Consumption habits

This factor is related to the customs and usage of a population for explaining the consumption habits and to know if a product or service can be accepted or refused in a domestic market. The habits of the population are very important in order to foresee whether a product can be popular or not to the population of a specific country.

Population growth:

This indicator calculates the population growth rate during a period of time. It is used to define the amount of target population to whom the product is targeted to.

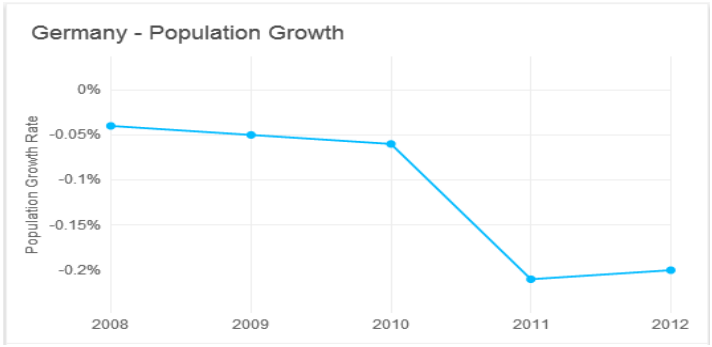


Figure 35: Population growth

Education level

This indicator measures education level of a population from the primary or elementary school to high level of education and learning experience. This is especially important when you need staff on arrival to provide specific services (e.g. post-sale, assembly, service, etc).

5.1.4 Technological

Nowadays the technology plays a significant role in the modern business, in order to be aware of the new technology of the other country we have to consider the following indicators.

Internet users

This indicator refers to the ratio of people with internet access over the total population.

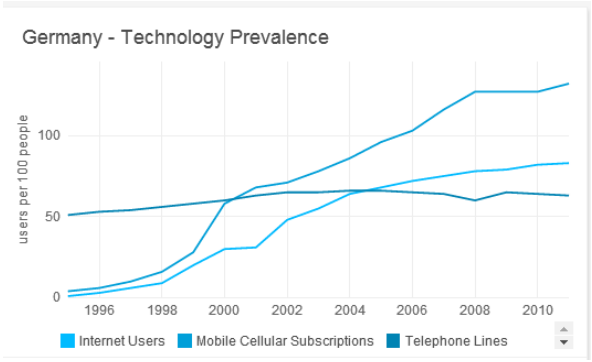


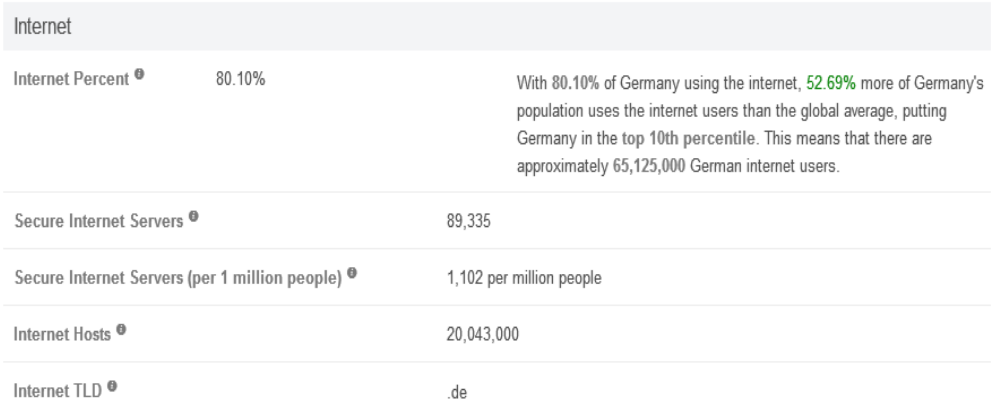
Figure 36: Internet users

Online shopping habits

This indicator is important in order to be aware of the consumption habits of the people.

Security of Internet Servers:

This indicator attempts to measure how safe the internet transactions are considered.



Internet		
Internet Percent	80.10%	With 80.10% of Germany using the internet, 52.69% more of Germany's population uses the internet users than the global average, putting Germany in the top 10th percentile. This means that there are approximately 65,125,000 German internet users.
Secure Internet Servers	89,335	
Secure Internet Servers (per 1 million people)	1,102 per million people	
Internet Hosts	20,043,000	
Internet TLD	.de	

Figure 38: Security of internet servers

5.1.5 **Legal**

Depending on the country that an SME is interested in exporting, the company might need to change or adapt the product by considering the legal dimensions. In some cases the legal implications could be very important and comprise a blocking factor for internationalisation. The following indicators are considered.

Consumer protection laws

This indicator is important in order to ensure the right of the consumers and also the regulations of competition and trade. Companies need to be aware of these laws in order to be protected.

Regulations

The regulations are important so the company is aware if they need to change and adapt something about their product.

Employment laws

This information is important for the SMEs in order to know the legislation regarding their potential employees.

Anti-trust law

This indicator refers to the knowledge about the law for the business practices on monopolistic and the interstate commerce.

6 EXTRACTION OF INDICATORS IN MULTISENSOR

This section aims at proposing sources and mathematical models for extracting and calculating the proposed indicators. In most cases we will rely upon existing definitions and metrics that are currently used. The proposed mathematical models will mainly focus in the areas that are relevant to the research performed in the project and specifically in the recently emerging information from social media.

6.1 Data resources and services for calculating the indicators

This section describes the type of resources based on which the proposed indicators can be calculated. To this end the following resources are considered:

- Extraction from web portals using web scraping tools
- Retrieved with APIs
- By performing content analysis (e.g. Name entity, concept, context extraction)
- Using the contributor analysis Module
- Using a sociometrics crawler
- Using widgets
- Using mathematical formulas that rely on information that can be extracted following the resources above.

A detailed empirical analysis of the first two (APIs and web portals) is also provided in the empirical study deliverable (D2.1).

6.2 Development of mathematical models

In this section we discuss mathematical models that can be used to model and forecast some of the indicators outlined in the previous sections. As already discussed these models will mainly focus in social media, however additional discussion on financial-relevant indicators is provided.

6.2.1 Modelling of social media monitoring indicators

The sharing of content through social networks has become an important mechanism by which people discover information, user-generated content, products and services. Social network platforms provide sharing and reposting functionalities that facilitate the diffusion of information through the network, by enabling users to simultaneously share information with their social peers and triggering a cascade of adoption.

Among all social indicators, we focus on the problem of predicting the popularity dynamics of content on online social media. Understanding the dynamics of online popularity (Ratkiewicz et al., 2010) has received considerable attention by many research communities due to its sociological and economic implications, as well as for important applications in marketing.

Social networking platforms offer the ideal setting where to study these phenomena. Prior work has focused on predicting the number of votes on Digg stories (Szabo et al., 2010) the

popularity of hashtags (Ma et al., 2013), messages (Hong et al, 2011), photo reshares (Cheng et al., 2014).

One of the most used models to forecast future popularity exploits the strong correlation between logarithms of the popularity at different stages. This can be expressed in the following form:

$$\ln N(t) = \ln (r_{t,t_0} * N(t_o)) + \epsilon(t, t_0)$$

where:

- N(t) represent the popularity at time (t)
- r_{t,t_0} Accounts for the linear relationship between the popularity at t and the popularity recorded at time t_o (e.g. after the first day)
- ϵ is a noise term.

6.2.2 **Modelling of financial indicators**

Most of the indicators proposed for SME internationalisation, e.g. GDP, can be modelled as macro-economic time series (Granger and Newbold, 2014). Thus, the task of forecasting such indicators can rely upon methods for the analysis of multivariate time series, which have been developed and tested during the last 30 years in the statistical-economics literature. Customers' intentions of buying a product depend of various factors, properties of the product, costs and the presence of competitors in the market. When entering a new market, it is important to analyse those factors and customize the product to meet user's needs. Since the development of mathematical models to predict financial information going beyond the scope of MULTISENSOR, in the project we will use predictions and estimations provided by external resources.

In the following section we summarise the indicators and provide specific resources and methods for their calculation.

7 CALCULATION AND SUMMARY OF THE INDICATORS

This section summarises the indicators introduced in the previous sections and suggests specific resources and methods for extracting and/or calculating them. The proposed resources are also based on the empirical study performed in D2.1.

7.1.1 Journalism indicators

Indicator	Description	Model/type/computation formula (if applicable)	Data acquisition (sources, APIs etc.)	Research task
Number of mentions by other media outlets	An indicator of the coverage of a topic by other media outlets.	A **number** corresponding to the total mentions of topic in media outlets, which should be specified as search/filtering criteria.	Can be collected via **APIs, e.g., Twitter** for content in social media or can be retrieved by a search query in the **central news repository** specifying media sources.	Identification of topics / information
Number of mentions by news agencies / independent correspondents	An indicator of the coverage of a topic by news agencies / independent correspondents.	A **number** corresponding to the total mentions of topic by news agencies/independent correspondents, which should be specified as search/filtering criteria.	Same as above.	
Number of mentions by main influential sources with regard to a specific topic (so-called „news hounds“)	An indicator of the coverage of a topic by main influential sources.	A **number** corresponding to the total mentions of topic by main influential sources. Influential users can be derived by exploiting the Contributor Analysis module which collects categories of interest for each influencer and computes influence scores for each user. Number of mentions returned can be controlled to include only mentions from users with specific interests that also have high influence ranking. The set IFL of influential users can be computed as IFL = set of top 5% influential users based on influence scores (see below for different influential scores) Q = {influential users relevant to topic W}	Same as above.	
Number of related hashtags	An indicator of the spread of a topic based on the number of related hashtags.	A **number** corresponding to the total related hashtags to the topic of interest. Similar to above for mentions, but instead, applied to hashtags. Potential noisy hashtags may need to be removed manually or using a frequency threshold.	Can be collected via Twitter API based on a specific search query.	
Position in the trending topics lists	Rank computed from list of trending topics, retrieved from API (e.g., on Twitter)	A **rank** corresponding to the position of a hashtag within the trending list provided by Twitter.	Can be collected via Twitter trending lists.	
Social media interactions	Number of tweets/retweets, likes, shares etc.	A **number** or **ratio** corresponding to total number of interactions caused by a user or by an activity, i.e., a post or a tweet	Depending on access rights can be computed exploiting the **related APIs**	

			of Twitter or Facebook.	
Patterns or anomalies (e.g. in big data analysis)	Patterns or anomalies found by the journalist revealing a new emerging topic or event.	Anomaly models can be executed on top of time series data to detect emerging topics. This is **not a directly measurable indicator** but requires the participation of the journalist.	Can exploit Twitter trending topics or other available tools for Twitter analytics.	
The content/information itself	The content item itself in the form of web pages, social media posts etc.	This is **not a directly measurable indicator**.	Content items from all media sources as found in the **Central News Repository** or other sources exploited independently by the journalist.	Relevance check
Keyword/entity/URL check	Check whether specific keywords, named entities, URLs etc. appear within the content items retrieved.	This is **not a directly measurable indicator**.		
Public discussion	Check whether there are discussion threads about the topic of interest. Extended public discussion would imply relevance for the general public.	This is **not a directly measurable indicator**.		
Related traffic	Check the traffic related to the topic of interest.	This is **not a directly measurable indicator**. Indicators such as total social media interactions and the position in the trending topics lists can be exploited for this purpose.	n/a	
Names	The names co-occurring with the topic of interest.	The named-entity recognition component identifies the **set of named entities** appearing in the search result.	Set of content items returned after querying the **Central News Repository** regarding the topic of interest.	Who?
URLs	The URLs co-occurring with the topic of interest.	A dedicated tool can support the extraction of **the set of URLs** appearing in the search result.		
Profiles (social networks)	The social profiles appearing with the topic of interest.	A dedicated tool can support the extraction of **the set of social profiles** appearing in the search result.		
Location	The locations appearing with the topic of interest.	The context extraction module associates **a location** with each content item. Additional locations can be found by manually inspecting the search result.		
Images	The images appearing with the topic of interest.	**A set of images** from the search result.		
Textual description	A description of the topic of interest.	The summarisation module is used to provide **a summary** (abstractive or extractive) from the set of content items returned by the search of the journalist.		What?
Images & videos	Images & videos appearing with the topic of interest.	**A set of images & videos** from the search result.		
Images & videos analysis	An analysis of the images and videos from the journalist to identify details regarding the topic of interest or certify authenticity (e.g., is the content genuine or has it been manipulated?)	This is **not a directly measurable indicator**, requires participation of the journalist.		

Reports by others	The reports by police and other state authorities, witnesses, media outlets.	This is **not a directly measurable indicator**, requires participation of the journalist identifying and telling apart such reports focusing on "what".		
Duplication	Multiple reports about the same event.	The use of the **summarisation module** can assist into removing duplicate information.		
Timestamp	Timestamps appearing with the topic of interest.	The context extraction module associates **a time element** with each content item. Additional timestamps can be found by manually inspecting the search result.		When?
Context	Contextual elements associated with the topic of interest.	This is **not a directly measurable indicator. Contextual descriptors** such as location, formality, time etc. can be exploited by the journalist to assist in assessing the context.		
Content metadata	Metadata associated with the topic of interest.	This is **not a directly measurable indicator.** Metadata in cases are accompanying the content items.		
Geo location	The geo locations appearing with the topic of interest.	The context extraction module associates **a location** with each content item. Additional locations can be found by manually inspecting the search result.		Where?
Implied location	Implied locations appearing with the topic of interest.	Same as above.		
Content metadata	Metadata associated with the topic of interest.	This is **not a directly measurable indicator.** Metadata in cases are accompanying the content items.		
Timeline (chronology)	A timeline associated with the topic of interest.	This is **not a directly measurable indicator**, requires participation of the journalist and/or sophisticated tools for identifying and ordering events associated with the topic of interest. The **event detection module** can assist in this process.		How?
Images/videos	Images & videos appearing with the topic of interest.	**A set of images & videos** from the search result.		
Reports by others	The reports by police and other state authorities, witnesses, media outlets.	This is **not a directly measurable indicator**, requires participation of the journalist identifying and telling apart such reports focusing on "how".		
History	History of the topic of interest	This is **not a directly measurable indicator**, requires participation of the journalist.		Why?
Background and position of involved individuals	Background information about involved individuals.	This is **not a directly measurable indicator**, requires participation of the journalist.		
Interests/motives	Interests and motives of involved individuals.	This is **not a directly measurable indicator**, requires participation of the journalist.		

Table 6: Summary of journalism indicators

7.1.2 Commercial media monitoring indicators

Indicator	Description	Model/type/computation formula (if applicable)	Data acquisition (sources, APIs etc.)	Research task
Circulation	Information about the circulation, printed copies and number of visits per medium.	A **number** corresponding to the total copies/visits for printed/online media.	Provided by the **publishers** themselves and fed into the repository by **PressRelations**. Part of the metadata of every print/online news.	Benchmark & Presence analyses
Reach	Estimation of readers consuming a medium.	Reach can be calculated on the fly as a **number** from the formula *Circulation x factor* or retrieved by rating agencies.	Media factors used in the computation of reach or reach directly can be acquired by **expert distributors** or **rating agencies**.	Benchmark & Presence analyses
Article numbers	Absolute number of articles w.r.t. a company, person or product defined as input.	Iterative search with respect to specified criteria returns a set of articles, for which its size as a **number** is reported across different time periods.	News items from all media sources as found in the **Central News Repository**.	Benchmark & Presence analyses
Type of media/article	Different types of media and articles allowing a more fine-grained analysis of the article numbers.	For media, type will be retrieved from an available list, while from articles the **context extraction module** will be exploited either extracting the type from the metadata or inferring it.	For media, **available lists** containing each media and its type. For articles, a **predefined list** of interesting categories such as interview, press release etc.	Benchmark & Presence analyses
Share of voice	Percentage of news items about a topic.	A **percentage** of matching items given criteria. Can be calculated by allowing refined search with formula *# object mentions / total mentions*	News items from all media sources as found in the **Central News Repository**.	Issues analysis
AVE (Advertising Value Equivalency)	Captures the cost/value of an advertisement.	A **number** corresponding to a financial cost.	Part of metadata of every print/online news fed into the repository by **PressRelations**.	AVE analysis
Initiative quotient/ratio	Captures the success of a company's public relations	A **ratio** capturing the articles initiated by a PR activity of the company manually computed by examination per article.	Every article in the set possibly retrieved from the Central News Repository needs to be coded for self- or externally initiated, done **manually**.	Presence analyses
Acceptance level/tonality/tone of voice	The sentiment of media items referring to specific products/companies etc.	A **number** computed as $\sum$ *tonality values / # measured items* where tonality can be computed by the **sentiment analysis module**.	News items from all media sources as found in the **Central News Repository**.	Presence analyses
Exclusivity	The centrality of a news item w.r.t. an object.	A news item can be either an **exclusive report, a focus report or a marginal report** (3 different options).	Manually acquired through examination by the PR expert.	Benchmark & Presence analyses
Author analysis	Information about an author.	A collection of available and inferred **metadata** about an author acquired automatically by the **context extraction module** and **social media analysis** module in the case of twitter.	Automatically acquired by the **context extraction module** and **social media analysis** modules using as input the content of the Central News Repository.	Stakeholder analysis
Picture Penetration	Captures media coverage of products/companies etc.	A metric capturing the **frequency** of positive exposure of logos etc. in media. Logos can be identified automatically using the **multimedia concept detection module**.	This is done manually.	Reputation analysis

Follower/following ratio	Capturing the influence of a user	A **ratio** computed as *# Followers / # Following*	Retrieved using available **Twitter APIs** per user.	Social media analysis
Interaction rate for Twitter	Capturing how often a user's tweet causes another user's interaction.	A **ratio** computed as *# Retweets of a Tweet + # Favorites of a Tweet / # Followers of the user*	Retrieved using available **Twitter APIs** per user.	
Interaction rate for Facebook	Capturing how often a Facebook post causes interaction.	A **ratio** computed as *# Likes + # Comments + # Shares / # Followers*	Depending on the page access rights, this can be calculated using **Facebook APIs**.	
User activity	Capturing the activity of a user	A **ratio** computed as *# Posts / {period}* for a specific period	Computed for public posts of user using **Facebook APIs**.	
Total interaction rate	Capturing the interaction across social media.	A **ratio** computed as *[TWITTER]((Favorites + Retweets) / Follower) + [FACEBOOK]((Likes + Comments + Shares) / Likes|Friends) + [YOUTUBE](Comments + Likes + Dislikes) / Views) + [BLOGS]((Comments + Social Shares) / Backlinks) + …*	Depending on access rights can be computed exploiting the **related APIs of Twitter, Facebook and Youtube.**	

Table 7: Summary of commercial media monitoring indicators

7.1.3 SME internationalisation indicators

Indicator	Description	Model/type/computation formula (if applicable)	Data acquisition (sources, APIs etc.)	Research task
Government type	The type of the government of a specific country.	The **government type** can be extracted from Wikipedia pages or other manually curated databases. A more reasonable approach would be to just update the DB once manually.	The Wikipedia API cannot return specific details such as the government type of a country but one should crawl/scrape that information. Alternative source: https://www.cia.gov/	Political analysis
Political instability	The political instability of a specific country.	An **index** of instability extracted from online sources.	Political Instability Index: http://viewswire.eiu.com/site_info.asp?info_name=social_unrest_table	
Corruption perception Index level	A metric of how corrupted is a public sector of a country.	An **index** of corruption extracted from online sources.	CORRUPTION PERCEPTIONS INDEX: http://www.transparency.org/cpi2014/results	
GDP Country	A metric of the economic power of a country.	The **GDP metric** extracted from online sources which provide a plethora of content in various formats (e.g., excel, csv) available to download, and also widgets that contain data displayed as tables, maps or charts, which can be embedded on a website. The widgets update automatically. The site also provides API access for creating custom data visualisations. **Best approach would be to include these widgets.** This can be modeled as time series. Prediction and trend-analysis can be performed if macroeconomic indicators are available. Projected GDP estimates are publicly available in Wikipedia: http://en.wikipedia.org/wiki/List_of_countries_by_projected_GDP_%28P	The World Bank data: http://data.worldbank.org/country/germany	Economic analysis

		PP%29_estimates		
GDP purchasing power parity (PPP)	A metric of the purchasing power of the population of a country.	Same as above.	The World Bank data: http://databank.worldbank.org/data/views/reports/tableview.aspx?isshared=true	
Inflation, consumer prices	A metric of inflation and consumer prices.	**Inflation** can be predicted by applying macroeconomic forecasting techniques. Predictive models are available[12]	The World Bank data: http://data.worldbank.org/indicator/FP.CPI.TOTL.ZG/countries/DE--XS?display=graph	
Customs and tariffs	Information regarding customs and tariffs such as taxes, VAT, fees, duties etc.	This is **not a directly measurable indicator**, information can be extracted from online sources. The link opens a .pdf document. Best approach would be to manually extract the needed information and store it in the DB or provide the link to the .pdf file.	Official Journal of the European Union: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:312:FULL&from=EN Alternative source (without API): http://ec.europa.eu/taxation_customs/dds2/taric/taric_consultation.jsp?Lang=en&SimDate=20150219	
Exchange rate	n/a	API for getting **exchange rates** is easily accessible.	Get exchange rates website: http://www.getexchangerates.com/api/	
Consumer Confidence	Measures the degree of optimism of the consumers.	**A set of metrics** available from Business and Consumer Surveys. The content is available in .pdf format, accessed through links. Suggested approach is to include the links in our DB.	Source: http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm	
Business Confidence	Measures the strength of businesses in a specific country.	Extracted from online sources, computed as *BCI(i) =[(∑(N,n=1)P(i,n)) /N] X 100*	Source: http://www.tradingeconomics.com/germany/business-confidence Paid access required.	
Export & Import	Volumes of exports and imports of services and goods per country.	Online source provides a plethora of content in various formats (e.g., excel, csv) available to download, and also widgets that contain data displayed as tables, maps or charts, which can be embedded on a website. The widgets update automatically. The site also provides API access for creating custom data visualisations. **Best approach would be to include the widgets.**	The World Bank data: http://databank.worldbank.org/data/views/reports/tableview.aspx?isshared=true	
Market share	The percentage that owns a company or business in a particular market.	Forecasting market share can be accomplished by statistical analysis of the market or of analogous markets and econometric methods. Can be computed as percentage as follows: *unit market share (%)=Unit Sales(#)/Total market unit sales (#)*	Provides API access; not sure if it's paid or free (requires to submit a request to the customer service). http://www.statista.com/topics/1739/yogurt/	
Picture Penetration;	n/a	This is **not a directly measurable indicator**, needs to be done manually.	n/a	

[12] http://www.mathworks.com/company/newsletters/articles/macroeconomic-modeling-and-inflation-rate-forecasting-at-the-reserve-bank-of-new-zealand.html

Competitors				
Country distance	The distance between two countries.	This is **a number in units of length** (kms) for which there are APIs such as the Google Distance Matrix API and the MapQuest Directions API that allow computing the geographic distances between cities. However, these services impose usage limitations that make it difficult to deliver a scalable solution, unless they are paid.	Google Distance Matrix API: https://developers.google.com/maps/documentation/distancematrix/ MapQuest Directions API: http://developer.mapquest.com/web/products/dev-services/directions-ws	

Table 8: Summary of SME internationalisation indicators

8 CONCLUSIONS

In this document, we have analysed the use of information and data from heterogeneous and multilingual sources by journalists, commercial media monitors and business managers. In this context we have identified the most important indicators for media monitoring case, journalist case and SME internationalisation. During the process of the indicator definition we started from conceptual indicators, which we tried to translate into measurable metrics. It appeared that in some cases some indicators where highly conceptual (especially in the journalistic use case) and a mapping of them onto a range of values was not possible. Towards the definition of measurable indicators we have also investigated the application of specific mathematical models. In this case we have mostly focused in deriving such metrics for indicators that are relevant to specific research activities of the project such as social media analysis. It was apparent that the definition of new formulas for financial, social and political indicators goes beyond the objectives of the project. To this end we come to a conclusion to reuse as much information available as possible by exploiting existing public data, web portals and APIs.

In the previous chapter we have provided a detailed summary of the indicators for each use case and how each of them can be extracted and/or calculated. These indicators will be used by the reasoning and the decision support modules in WP5 especially for assisting the decision making process in the SME internationalisation scenario. The indicators for media monitoring and journalism will also be considered for the derivation of important facts using the reasoning services (WP5).

Overall, the main contributions of this deliverable are a) the definition and the analysis of the indicators from the user perspective and b) the attempt to quantify and identify ways to extract and calculate these indicators.

9 REFERENCES

Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J. 2014 "Can cascades be predicted?", Proceedings of the 23rd international conference on World wide web, p. 925–936.

Granger, CWJ., and Newbold, P. 2014. "Forecasting economic time series", Academic Press.

Hong, L., Dan, O., Davison, B.D. 2011. "Predicting popular messages in twitter", Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11, pp.57–58.

Jeffries-Fox, B. 2003. "Advertising Value Equivalency", White Paper for the Institute for Public Relations. URL: http://www.instituteforpr.org/wp-content/uploads/2003_AVE1.pdf

Ma, Z., Sun, A., Cong, G. 2013. "On predicting the popularity of newly emerging hashtags intwitter", Journal of the American Society for Information Science and Technology 64(7), 1399–1410.

Ratkiewicz, J., et al. 2010. "Characterizing and modeling the dynamics of online popularity", Physical review letters 105.15 (2010): 158701.

Szabo, G. and Huberman, B. 2010. "Predicting the popularity of online content", Commun. ACM 53, 8 (August 2010), 80-88.