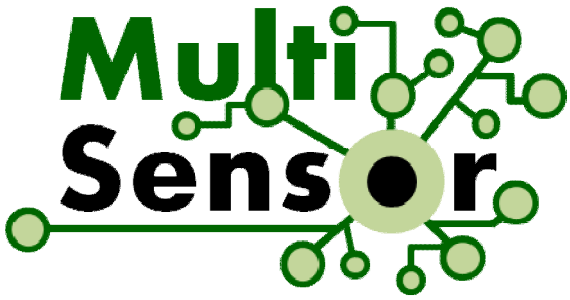


MULTISENSOR

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

FP7-610411

D9.4

Market analysis and initial exploitation plan

Dissemination level:	Public
Contractual date of delivery:	Month 18, 30 April 2015
Actual date of delivery:	Month 18, 30 April 2015
Work package:	WP9 Dissemination and Exploitation
Task:	T9.4 Exploitation plans
Type:	Report
Approval Status:	Final Draft
Version:	2.0
Number of pages:	38
Filename:	D9.4_InitialExploitationPlan_2015-04-29_v2.0.pdf

Abstract

This document is the limited public version of the initial exploitation plan for MULTISENSOR project. It designed to ensure general usability and exploitability of the project results by illustrating how MULTISENSOR technologies and tools could contribute to the everyday tasks of journalists, media monitoring companies and export managers for SMEs which are on the brink of internationalization.

Co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	11-03-2015	Initial structure	Michael Jugov
0.2	03-04-2015	Contributions	ALL
0.3	08-04-2015	Integration of first round of contributions from partners	Michael Jugov
0.4	15-04-2015	Additional contributions	ALL
0.5	20-04-2015	Integrated version for internal review	Michael Jugov
0.6	20-04-2015	Internal reviews	Stefanos Vrochidis Reinhard Busch
0.7	22-04-2015	External review	Eric Karstens (Advisory Board member)
0.8	27-04-2015	Final round of contributions from partners based on review comments	ALL
1.0	29-04-2015	Final integrated document	Michael Jugov
2.0	30-04-2015	Final limited document for publication	Michael Jugov

Author list

Organisation	Name	Contact Information
pressrelations	Michael Jugov	michael.jugov@pressrelations.de
pressrelations	Mirja Eckhoff	mirja.eckhoff@pressrelations.de
pressrelations	Romina Gersuni	romina.gersuni@pressrelations.de
DW	Tilman Wagner	tilman.wagner@dw.de
DW	Nicolaus Heise	nicolaus.heise@dw.de
PIMEC	Teresa Forrellat	tforrellat@pimec.org
PIMEC	Marti Puigbo	mpuigbo@pimec.org
everis	Alan Mas Soro	alan.mas.soro@everis.com
everis	Axel Callabed Emperador	axel.callabed.emperador@everis.com
everis	Emmanuel Jamin	emmanuel.jean.jacques.jamin@everis.com
everis	Marcel Risques Andersen	marcel.risques.andersen@everis.com
BM-Y!	Ioannis Arapakis	arapakis@yahoo-inc.com
BM-Y!	Iris Miliaraki	irismili@yahoo-inc.com
Linguatec	Reinhard Busch	r.busch@linguatec.de
Linguatec	Vera Aleksic	v.aleksic@linguatec.de
Ontotext	Boyan Simeonov	boyan.simeonov@ontotext.com
Ontotext	Vladimir Alexiev	vladimir.alexiev@ontotext.com
UPF	Gerard Casamayor	gerard.casamayor@upf.edu
CERTH	Stefanos Vrochidis	stefanos@iti.gr

CERTH	Dimitrios Liparas	dliparas@iti.gr
CERTH	Ilias Gialampoukidis	heliasgj@iti.gr
CERTH	Ioannis Kompatsiaris	ikom@iti.gr

Executive Summary

This market analysis and initial exploitation plan has been prepared with the goal of raising the consortium's awareness for the exploitability of the researched and implemented technologies, tools and services, to provide a basis for internal discussions and to generally increase the chances of the exploitation of MULTISENSOR after the project's end by effectively guiding the execution of the next stages of MULTISENSOR.

Taking into account the fundamental differences of the journalistic scenario, the commercial media monitoring scenario and the business intelligence scenario we have first tried to analyse user needs, the current market situation and what kind of existing tools are already available for users of each scenario.

In the case of journalists, while multilingual and multimedia media research is undoubtedly a notable topic, the key downside lies in the fact that journalists tend to mistrust automatically generated or detected information and trends. Also, while journalists are already adept at using SaaS media monitoring solutions for their research needs, they are usually not accustomed to paying for these.

The major difficulty in targeting media monitoring companies for commercial exploitation lies in aligning the tools and services closely alongside already established processes and workflows. On the other hand, commercial media monitoring companies would be willing to pay for solutions that prove to increase efficiency and/or open market potential for up-selling new products and services to their customers.

The unique selling proposition and one of the strongest differentiation points that was uncovered in the SME Internationalization use case is the ability to quickly provide actionable business intelligence for export-oriented SMEs in the form of reports, recommendations and analytics. Here, on the other hand, the biggest risk could be total market size for such a highly specialized application.

Of course, in the current stage of the project no evaluation of the project results has been performed. It is expected that the evaluation plan will need to be adapted and updated in line with project results in order to exploit in a more efficient way and uncover areas of higher innovation potential from a user and market perspective. This update will be reported in the upcoming deliverable D9.7.

Abbreviations and Acronyms

ARD	Association of public service broadcasters in Germany
ASR	Automatic Speech Recognition
BMCO	Broadcast Mobile Convergence
DAML	DARPA Agent Markup Language
DID	Digital Item Definition
DII	Digital Item Identification
DRM	Digital Rights Management
EBU	European Broadcast Union
EC	European Commission
ETSI	European Telecommunications Standards Institute
IEEE	Institute of Electrical and Electronics Engineers
IP	Integrated Project
IPTC	International Press Telecommunications Council
IST	Information Society Technologies
JPEG	Joint Photographic Experts Group
MAF	Multimedia Application Format
MMC	Media Monitoring Company
MPEG	Moving Picture Experts Group
MT	Machine Translation
NER	Names Entity Recognition
NITF	News Industry Text Format
NoE	Network of Excellence
OWL	Ontology Web Language
OWL-QL	Ontology Web Language Query Language
OWL-DL	Ontology Web Language Description Language
RDF	Resource Definition Framework
RSS	Really Simple Syndication
SaaS	Software as a Service
STREP	Specific Targeted Research Projects
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
SWRL	Semantic Web Rule Language

Table of Contents

1 INTRODUCTION

D9.4 is a core deliverable designed to ensure general usability and exploitability of the project results. Its goal is to illustrate how MULTISENSOR results will be exploited in order to contribute in a day-to-day working environment for journalists, media monitoring companies and as a decision support system for business intelligence focused on internationalization.

Discussing commercial exploitation requires an understanding of the technical framework of MULTISENSOR, i.e. the requirements and interdependencies of a unified platform designed to allow for a multidimensional content integration from heterogeneous sensors, and a thorough analysis of these requirements alongside business perspectives and market conditions to show how potential investors can possibly benefit from a unified platform such as MULTISENSOR - or pieces thereof.

In D8.2 the three MULTISENSOR user partners i.e. Deutsche Welle, pressrelations and PIMEC have used their specific experience and expertise in order to define user requirements. The proposed three use cases reflect the common challenge of having to deal with a large amount of heterogeneous data and information from different sources in different languages. Thus, MULTISENSOR's common requirements are to help users:

- to understand the meaning of a specific content
- to provide decision support by clustering what belongs together and allowing users to quickly discard what is irrelevant and
- to analyse the relevant data

Yet, while journalists, media monitoring professionals and business executives alike require systems that streamline and analyse news content, we also found very significant differences:

- **Journalists** require a research platform that delivers continuous news data on changing topics from many different sources and of different types. The core requirement is to allow journalists to quickly capture a complete picture without missing valuable information and to be able to drill-down on specific findings by ad-hoc analysis.
- **Commercial Media Monitoring Companies** also need to continuously monitor specific topics, brands or campaigns and they also need to perform data analysis. Their set of criteria, always individually tailored to their customers' specifications, is usually more static than that of journalists. Through data curation they must clean the data of unwanted noise to deliver to customers only what fits. Data analysis is often performed within a hybrid framework (computer-aided decision support plus human validation). Data enrichment, such as summaries or translations of news articles or a network analysis of news contributors, relies largely on manual editorial effort and desk research.
- **Business Intelligence** companies and organizations need to collect, store and analyse market information such as brand strength and -awareness, activities by competitors, pricing, market barriers, legal restrictions and statutory requirements and to monitor important stakeholders from their respective domains. While such information is often found by monitoring international and social news content, there is an

additional requirement to evaluate those findings alongside relevant socio-economic indicators to allow for well-founded business decisions - such as whether a company should move into a new market or not.

When discussing exploitation strategies, these fundamental differences need to be taken into account - which is why they are reflected in this document's structure. Consequently, the journalistic scenario, the commercial media monitoring scenario and the business intelligence scenario are described and analysed in separate sections. Building upon a thorough market analysis on the three domains of interest for the project, in this initial exploitation plan we will try to develop ideas and plans for exploitation, based on specific demand and particular market conditions.

Overall, the D9.4 aims at providing a more detailed market analysis (compared to D8.2) and business models for the domains considered in MULTISENSOR i.e. journalism, media monitoring and SME internationalisation (section 2). The document specifies the exploitable foreground to arise from the project, outlines specifications of the technical modules (section 3) and provides the plans for exploitation by all the partners (section 4).

2 BUSINESS MODELLING

The central goal of MULTISENSOR is to develop tools that are of interest for three different business segments:

- Journalistic research
- Commercial media monitoring
- Business intelligence decision support

In D8.2 we have analysed user requirements with regard to the specific use case scenarios. Since the above scenarios are quite different from each other also from a perspective of exploitation, we have decided to again follow a more or less similar approach.

For each use case, we try to pinpoint the relevant market segment for the MULTISENSOR project results and overview the prevalent and foreseen general business models within that market segment. Built on the preliminary market survey, as well as interviews and discussions with relevant stakeholders performed by each user partner within their respective sub user groups, we analyse existing market solutions in regards to their range of offered services and pricing. This lays the foreground for exploitation within potential target markets that can benefit directly from the technologies and services developed within MULTISENSOR.

2.1 Journalistic Research

The field of journalism is in constant movement ever since the distribution and production processes have been overthrown by the developments of the digital age. This leads to many new requirements by media companies, looking for new ways to research and find news, detect trends and make sense of the large amounts of information that are nowadays available to everyone. Bringing new tools to this field is both challenging and rewarding at the same time. In order to spot opportunities we'll have a closer look at the current market situation and compare it to what MULTISENSOR has to offer to find possible access points for exploitation.

There are two particular areas, where MULTISENSOR may render the journalistic work more efficient, one being the area of news and the other the field of investigative journalism. Looking at the processes of news research, MULTISENSOR would streamline these tasks, by offering automatically generated summaries that help reduce the effort of pursuing a multitude of original sources manually. This would allow for a clearer and more nuanced picture of any news item more easily than today.

The investigative/original type of journalism, on the other hand, could use MULTISENSOR as a discovery tool and reference. This would help journalists to more easily find sources and opinions, even in other languages that are relevant to an investigation, and avoid being surprised by aspects of full stories already being published by others.

For a clearer view on these opportunities we'll have a closer look at the current market situation and compare it to what MULTISENSOR has to offer to find possible access points for exploitation.

2.1.1 Market Analysis

As part of the requirements gathering for the MULTISENSOR Project, DW conducted a thorough market analysis, looking at companies and tools for online research, media monitoring and analysis available on the market. For finding as many options as possible DW relied on three main channels:

- it made use of the expertise of its own journalistic network, interviewing its members about their preferences,
- it tapped into different social media channels and
- it conducted a general online research, focusing in both cases on the main research areas of the MULTISENSOR project.

Overall, the analysis clearly showed that there was no one tool covering all envisioned MULTISENSOR functionalities (see Figure 1). In order to still be able to compare the tools with each other and MULTISENSOR, they had to be categorized by their respective functionality focus. The categories covered *Social News Aggregators* (such as **virato**[1]), *Social Network Search and Analysis* (e.g. **topsy**[2]), different *online media monitoring services* (e.g. **Newsexplorer**[3]) as well as tools for *text analysis, extraction and comparison* (e.g. **Semantic Wire**[4]**)**

This allowed for a good overview of what functionalities are currently being offered to users on the market. It also made clear where MULTISENSOR could fit in and fill a gap in this landscape of media monitoring and analysis, both in the combination of the different analysis elements as well as in regards to simple elements such as a simple search screen.

	automatic language detection and translation	multiple source integration	analysis				entity detection	categorization	enrichement	automatic summarisation
			sentiment	semantic	text structure	networks				
Social News Aggregator	NO	twitter, facebook, G+, Blogs	NO	NO	NO	YES	NO	YES	NO	NO
Network and Search analysis	NO	twitter only	partly	NO	NO	partly	NO	partly	NO	NO
Online MM Services	interfaces only	social media & the web (to different extents)	NO	NO	NO	partly	partly	partly	mainly not	NO
Text analystics, extraction & comparison and web news filter	partly	mainly manual input necessary	mainly not	partly	partly	mainly not	mainly	mainly	NO	NO
MULTISENSOR	**WP2**	**WP2 & WP4**	**WP3**	**WP5**	**WP3**	**WP3**	**WP5**	**WP5**	**WP5**	**WP6**

[1] http://www.virato.de/

[2] http://topsy.com/

[3] http://emm.newsexplorer.eu/NewsExplorer/home/en/latest.html

[4] http://www.semanticwire.com/

Figure 1: Comparing planned MULTISENSOR functionality with tools currently available (short version. For more details check Deliverable D8.2)

The research also made clear what people are willing to pay for, as there was a large number of freely available tools and quite a number of paid-for offerings.

2.1.2 Business Definition

Most of the services for media monitoring available on the market today are Software as a Service (SaaS) Solutions – tools that allow access via a web interface and don't have to be installed or integrated into the systems of a news organisation.

This approach offers the best opportunities, as it allows for a broader customer base than a standalone software solution that has to be installed separately. Firstly, because it avoids the difficulties of integrating the product into a technical landscape that is unknown to the project beforehand. Secondly, because this allows for offering the solution also to SMEs or single customers, such as freelancers, who don't necessarily own the necessary hardware setup.

A very common option for this scenario is to offer several business-models targeting different user groups. A limited free version to start with helps attracting customers and allows for quick individual usage. This includes early-adopters who could share their experience and help promote the product. This model can then be adjusted and built upon. The second step would be a licence for small and medium sized companies or freelancers who are willing to pay for the use of the service, but only need one to a few user accounts. On top of that MULTISENSOR could offer a heavy user licence for companies with many employees, such as media companies.

Depending on the general state of the prototype and the actual demand on the market at the end of the project, it is conceivable to promote MULTISENSOR as a small media start-up, offering this kind of service to customers in the field of journalism and media.

2.1.3 General Market Conditions

Among the different factors that have to be taken into account before taking the step towards market entry is the current market situation.

At the moment, the market for media monitoring tools is in a constant state of flux. The market analysis has shown the existence of an already broad spectrum of different tools, covering a wide range of functionalities. All these tools are competing for the customer's attention, while new players emerge constantly, offering improved functionalities or new combinations of such.

Hence, one has to be either quick with a new idea and offer something unique that no one else has offered so far. Or one has to enter the market in a niche and offer a very comprehensive and well-designed product that covers one aspect – but that one better than anyone else.

Looking at the market this is no easy task, as one would face grown companies such as *"Meltwater"* on the one hand, but also compete with established and upcoming start-ups such as *"topsy"* or *"interpretive"* on the other hand. All of them offer products with lean interfaces, more or less easy to use. They all offer well-developed search algorithms,

different analysis models, ranging from very simple to more complex and lean and (more or less) easy to use interfaces.

Bringing a product to market such as MULTISENSOR, that allows for multilingual and multimedia search across several different sources, including both quick in-depth analysis support through summarisation, sentiment analysis, network analysis and concept detection could work, just as there doesn't seem to be anything as comprehensive in one single interface. One feature that is particularly valuable for the European market is the support of several languages, allowing for the in-depth analysis of news not just in your own language. With the EU constantly striving towards a common public sphere, but being hampered by its many different languages and cultures, this could mean a great step forward. Journalists would have access to news and information from around the EU available at their fingertips. They wouldn't be limited to the languages they speak and a predominant English working atmosphere, but could work with news from all member countries (as far as the languages are covered) alike. Of course this is not only limited to the EU but could also be very valuable in a market that is spanning across the globe – both in terms of sources as well as audiences.

2.2 Commercial Media Monitoring

Media monitoring, news aggregation and its distribution is a large and growing global market segment. Rapid development of digital technologies has led to a dramatic increase in news content worldwide and has lowered the cost of harvesting and storing it.

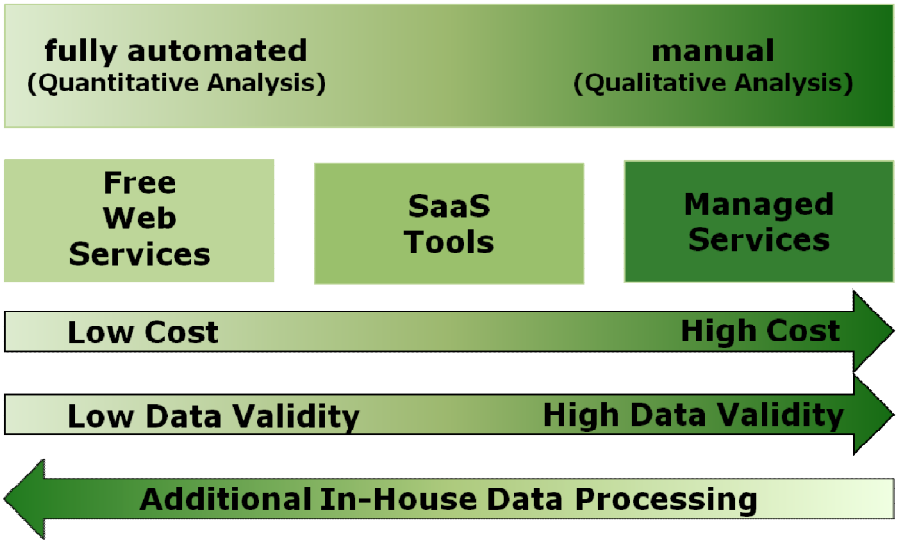
Our comments in regards to exploitation are based on thorough market analysis, pressrelations' own experience as a long-time vendor in the media monitoring industry as well as the expertise of a network of media monitoring companies and media monitoring experts which we have organized in a media monitoring sub user group. The idea behind such a sub user group was to pinpoint the most promising areas of potential exploitation while trying to avoid being "business-blind" by seeing only one's own particular requirements, needs and "wish-list items". The sub user group consists of top-level executives with whom we have conducted a small number of in-depth interviews and discussions about their primary requirements and needs.

2.2.1 Market Analysis

In recent years, software development in the field of media monitoring has been with a strong emphasis on creating high-tech *Software-as-a-Service* (SaaS) platforms designed to support marketing- and media relations automation. Today's market is populated with inexpensive (even free) SaaS solutions. In contrast to these, there are media monitoring companies who offer *managed services* (driven by a so-called hybrid production workflow which combines data, technology and human data curation and analysis).

Commercial media monitoring companies on the SaaS- or automation-side of the continuum usually focus only on parts of the media spectrum (for example only on online and social or only on TV-monitoring) – mostly because such content is inexpensive to harvest – while managed media monitoring service providers usually offer a full 360° media spectrum (print, broadcast, online and social). Such systems, although generally more expensive for the customer, allow for a higher degree of accuracy, better customization and a near-perfect and customer-specific dissemination of results.

Generally speaking, the SaaS-group of media monitoring companies offer their services to customers who opt to *in-source* their media monitoring operations while managed services



by definition target customers wishing to outsource such services.

Figure 2: Different Systems for Media Monitoring

2.2.2 Business Definition

Instead of looking at the commercial business of media monitoring from the perspective of a highly automated SaaS-Platform, we will discuss potential exploitation of MULTISENSOR from the perspective of services targeted towards a managed media monitoring service providers. We chose this view both because this is pressrelations' market position, but also because the potential for exploitation as a pure SaaS tool is discussed within the journalistic as well as the SME internationalization use case.

In exploring exploitation of MULTISENSOR from the perspective of a hybrid media monitoring workflow, we can identify and segment the entire media monitoring production chain into:

- capturing and harvesting news content
- curating and analysing news content and
- delivery and dissemination of news content to customers

This media monitoring workflow is supported by various technologies:

- **Capturing and harvesting technologies** are designed to deliver print media, broadcast media, online media and social media into news repositories, from where they are made available for further processing and archived in line with national copyright regulations. Particular harvesting and capturing technologies exist for each media type and solutions to process the different media types and technologies and tools are usually laboriously assembled and collaged by MMCs from different vendors and self-developed modules.
- **Technologies for data curation and data analysis** are deployed on top of live news repositories and they are designed to help editors and analysts process huge amounts of daily news content. Normally, *editors* perform simple data curation tasks

while *analysts* go beyond selecting and filtering by extracting relevant information, measuring and scoring news items. This involves assessing topics, finding key messages, evaluating sentiment etc. Data curation and data analysis is normally performed in a time-critical environment which makes it essential that the technological system of choice presents news content in an intelligent and structured way, using suitable visualisation techniques which allow editors and analysts to grasp the necessary information at little more than a glance so as to perform their tasks quickly and efficiently.

- **Technologies for news delivery and dissemination** to customers are purpose-built according to the specific requirements of customers normally looking at news primarily through media relations glasses. They include features such as automatic generation and dissemination of media reviews (in various formats), news alerts (in case of an unusual level of media activity or adverse events) and comprehensive dashboards (online portals) showing aggregated and enriched analysis results.

Leaving aside the option to evolve MULTISENSOR into a full commercial media monitoring operation, a possible exploitation strategy may be geared towards providing technical services to established media monitoring companies. In this regard, while exploitation could potentially target all three segments of the media monitoring process chain, the most promising area of exploitation is *news data curation and analysis*. Here, we can observe a noticeable lack of marketed tools and technical solutions for efficiently processing an ever-growing load of news content within a hybrid workflow and it is here, where we see the best potential for commercial exploitation of MULTISENSOR results.

2.2.3 General Market Conditions

A competitive market analysis of the media monitoring market is designed to yield:

- **Overall market size** by evaluating available market statistics and surveys from the biggest associations in the sector **AMEC**[5] and **FIBEP**[6].
- **Available technological solutions** by focusing on active technology providers servicing media monitoring companies, with a special attention given to innovative technological solutions.

The size and makeup of the global media monitoring market can be deferred from the latest FIBEP State of the Industry Report 2013-14. The below chart illustrates the split of FIBEP members companies by their 2013 revenue. The total revenue of all 74 companies that provided figures (of 84 surveyed companies) was EUR 803.500.000 in 2013:

[5] International Association for Measurement and Evaluation of Communication

[6] Fédération Internationale des Bureaux d'Extraits de Presse

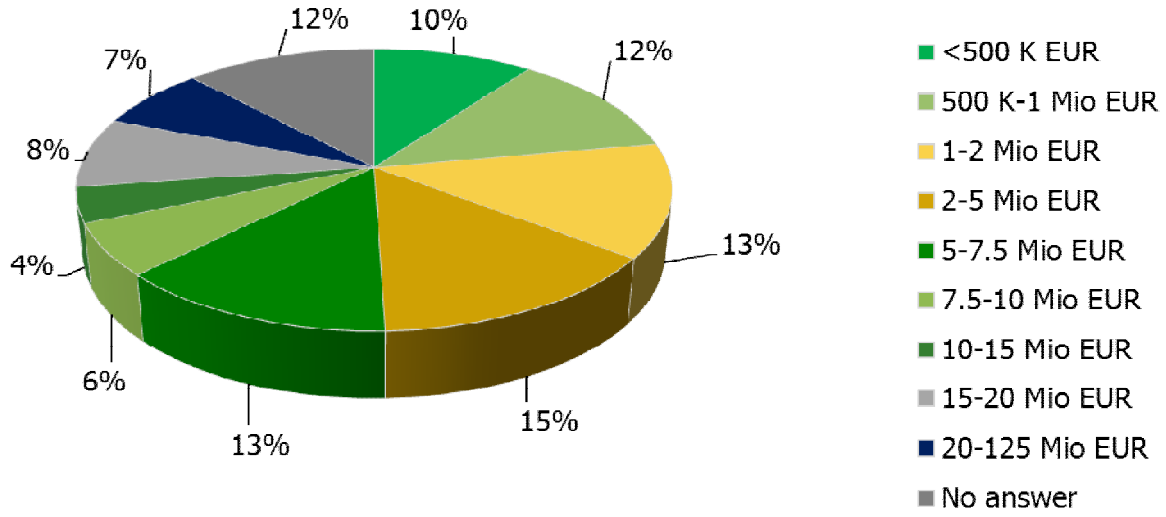


Figure 3: 2013 Revenue of Media Monitoring Companies organized in FIBEP

Media monitoring companies have seen a widening in their portfolio of services offered to customers in line with growing demands in a digital media age. The fastest growing demand has been to deliver Social Media Monitoring as well as Web News Monitoring.

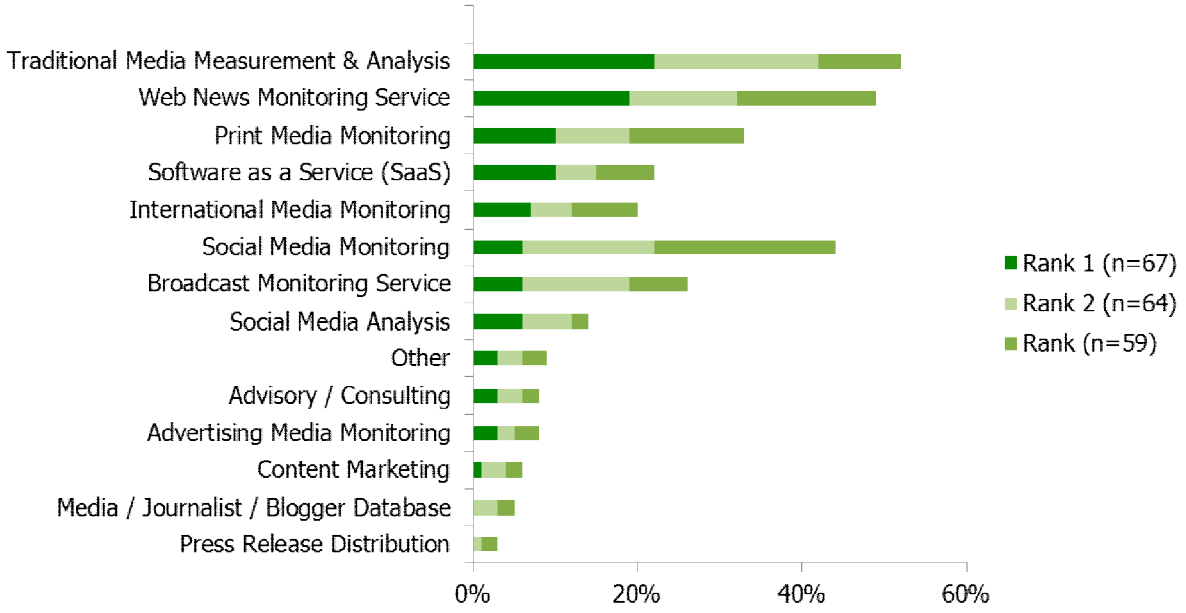


Figure 4: Services offered in 2013 by FIBEP members

In regards to multilingual services offered by media monitoring companies, the survey confirms our presumption that media monitoring is to a large extent still organised in the form of regional businesses. Traditionally, a media monitoring companies' competitive advantage has been to offer comprehensive nationwide coverage of their national market in regards to all or most local print titles and broadcasting channels.

Only in recent years – mainly due to the availability of inexpensively harvested online and social media news content – have national media monitoring companies been confronted

with having to deal with multilingual content. One of the answers to this challenge has been the recent advent of multinational media monitoring conglomerates (Cision[7] in the US, Kantar Media[8] in Europe and iSentia[9] in the Asia-Pacific region leading the drive towards internationalisation).

This below chart illustrates language coverage of the five MULTISENSOR languages (English, German, French, Spanish and Bulgarian) according to the FIBEP survey:

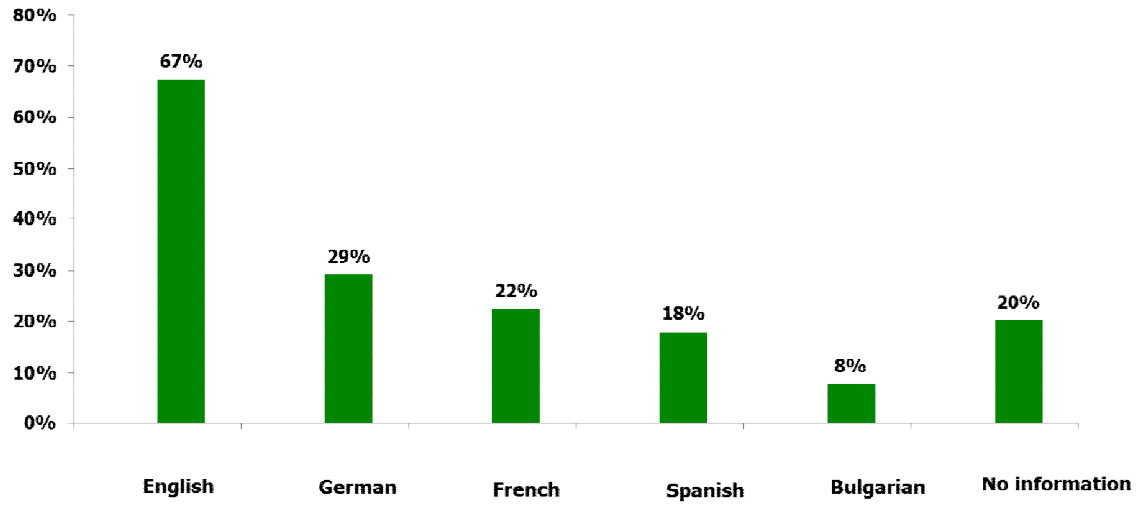


Figure 5: Percentage of media monitoring companies offering foreign language services

Eye-catching is the fact that one third of all surveyed companies are not able to process English language news content. This points to just how regionally-focused some media monitoring companies still work. News items from the other major European languages (German, French and Spanish) can only be processed by a small minority of surveyed companies.

[7] http://www.cision.com/

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

[9] http://www.isentia.com/

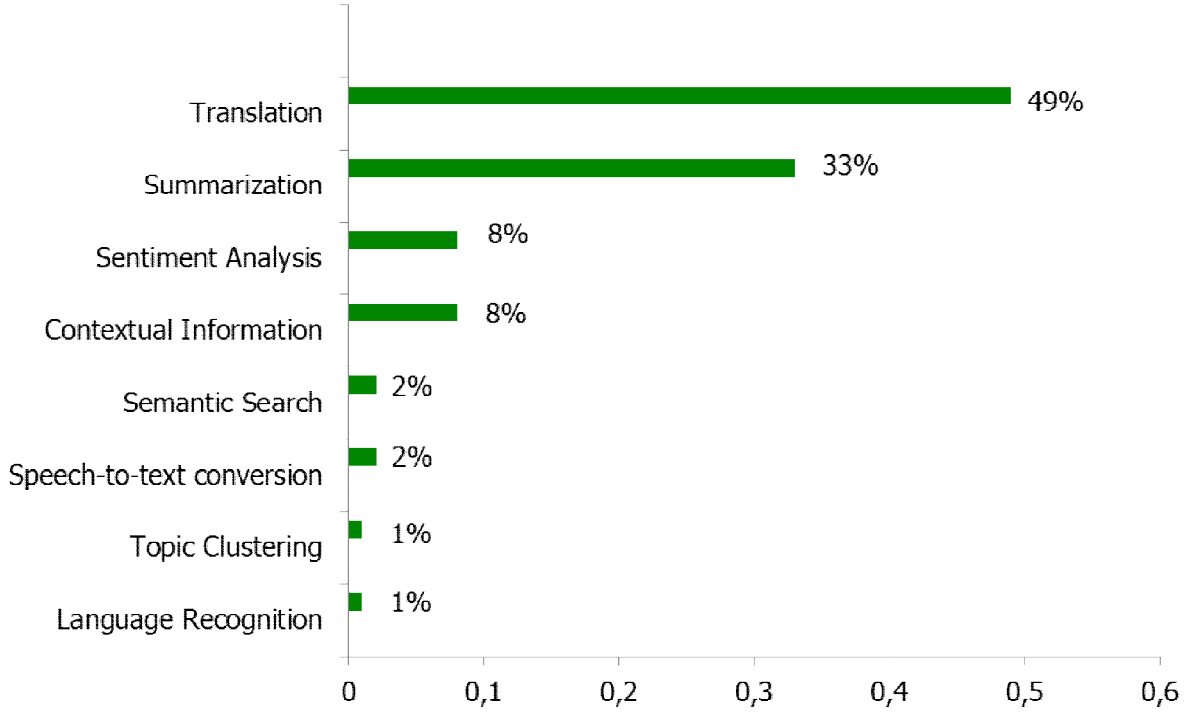


Figure 6: Technologies deployed at media monitoring companies

Asked about their mix of technologies the recent survey reveals quite a low level of deployed technological support and a conversely high level of manual editing and data analysis (the comparatively high level of technological translation services very likely stemming from copy-and-pasting text into **Google Translate**[10] and similar (free) translation services).

If MULTISENSOR is to uncover its potential as a *technology provider* for commercial media monitoring companies, exploitability will require a technical infrastructure through which MULTISENSOR services can be called by proprietary internal systems run by media monitoring companies. From the perspective of possible exploitation and given the limited scope and timeframe of the MULTISENSOR project we recommend to focus on the most important requirements, i.e. the ones with the highest exploitative value - i.e. summarisation, translation, and social media contextualisation as previously stated. Some thoughts may need to be given to how a module can run by itself by freeing it from the dependencies from other modules that may not be necessary or expedient for exploitation.

2.3 Business Intelligence Decision Support

2.3.1 Market Analysis

The market analysis for this pilot use case (PUC3) has been carried out in collaboration with SMEs and export managers. SMEs, despite the fact that they might not directly study the market, they have the final decision, and it is important to take them into account throughout the whole process. Export managers are the ones that directly manage the process of internationalisation of a company. On their daily basis they are looking for

[10] https://translate.google.com/

information and helping SMEs to export their products, so they have a broad knowledge about opening new markets and the actions to take in order to make decisions. For this matter, we worked with a group of freelance export managers, some of whom collaborated in joint projects with PIMEC.

Together with these specialists, we have confirmed that there are only a few portals and tools that are helpful for SMEs that intend to expand in new markets. National institutions such as the Spanish ministry of commerce or international/global institutions like the World Bank only offer very preliminary and basic information about the targeted country. It is also possible to find some basic and sometimes sector-related information in English that is provided by the targeted country itself, like through the German Business Portal iXPOS[11]. In addition, the '**Your Europe, your opportunity**[12]' portal offers mainly legal information on product requirements, competition and public contracts, among others. It has value on offering specifics on regulations and EU funding programmes.

But when digging deeper and looking for the crucial bits of information, it is hard to find anything that is available in Spanish or in English. This refers to up-to-date market data, consumer statistics as well as to very specific pieces of information like relevant legislation or restrictions. Looking at the comparison of these few tools with the planned functionalities in MULTISENSOR, but also the lack of more tools clearly shows the relevance of the MULTISENSOR project.

2.3.2 Business Definition

There are two compatible business strategies and targets to try and make MULTISENSOR evolve into a commercially viable platform. First, to offer the media analysis, country indicators and economic sector information to individual SMEs so they can assess it and take informed decisions on their internationalisation. A tailored and more narrowed-down service can be offered to them delimiting the access to the sector or sectors they are interested in, so that the proposals can be flexible and adapted to the needs and possibilities of each SME. Second, to offer the tool to entities such as chambers of commerce, export managers, etc. A complete access to the platform can be very helpful for these entities or managers that help SMEs on their export strategies and internationalisation processes. MULTISENSOR can offer refined information for entities that deliver assessment services to companies. The platform can strongly reduce the information costs and allow the councillors to offer their advice and export business strategies to a major number of clients and in a more informed manner.

In all, export managers at individual SMEs, as well as entities who consult and support SMEs such as chambers of commerce are the two business segments that MULTISENSOR can be targeted as potential user-clients. A Software-as-a-Service for internationalisation decision-making support can be offered in both cases, with differentiation on the commercialisation schemes. MULTISENSOR can be exploited in tailored premium schemes for individual SMEs and heavier premium licences for consultancies who use the service for multiple clients.

[11] http://www.ixpos.de/

[12] http://europa.eu/youreurope/business/index_en.htm

2.3.3 General Market Conditions

The commercially exploitable feature that MULTISENSOR can offer is the complete and refined information package of the socioeconomic situation of a country, including its market trends and consumption habits, and a specific sector within it.

MULTISENSOR delivers summarisation, it captures media sentiment and it gathers indicators data. Moreover, it can offer it in multiple languages and produce country comparisons with a focus on specific economic sectors. For that matter, the platform has an added value compared to other online tools such as **FindTheBest**[13] and **Indexmundi**[14], where you can check and compare multiple indicators. FindTheBest offers information, rankings and comparison tools in a variety of topics such as transportation, real estate, electronics or countries. Similarly, Indexmundi encapsulates many World Bank indicators and gathers them together in order to visualize them in an organised and efficient manner. In both cases, the access to their information is free but those tools do not go beyond the mere recollection of data indicators. MULTISENSOR, on top of that, delivers multidimensional content information from heterogeneous sources in a unified platform.

A service that constitutes a closer competitor to the MULTISENSOR, although with characteristics that differ, is the **Santander Trade**[15] portal. It is a portal that specifically targets companies that want to start their exportation activity or internationalisation. Santander Trade offers information that refers particularly to the needs of internationalisation, that is, market trend analysis, business counterparts records, shipments information and currency data. The service is offered to Banco Santander clients, which constitutes a barrier, and presents some paying features. For instance, more precise information on market trends, access to financial reports or to standards regulation require a fee to make them available. As a commercial strategy, Santander Trade offers a 30-day trial on their services and a demo access for the case of China in order for the clients to test and consider the benefits of their platform. They also try to establish synergies between companies that are clients of the bank and perform their activities in the countries where the bank operates. Last, the site has a global scope and covers information of all the countries, not only Europe.

[13] http://www.findthedata.com/

[14] http://www.indexmundi.com/

[15] https://en.santandertrade.com/

3 DEPLOYED SERVICES AND OPEN SOURCE PRODUCTS

The following table lists the initial exploitation prospects for the modules developed in MULTISENSOR.

#	Outcome	Deliv.	Exploitation prospects
1	Name Entity extraction tool	D2.4	Exploitation by LT
2	Concept extraction module	D2.4	Freely available
3	Multimedia concept extraction framework	D2.4	Freely available
4	Machine translation module	D2.4	Exploitation by LT
5	Context analysis module	D3.4	Exploitation by BM-Y!
6	Sentiment extraction module	D3.4	Freely available
7	Social media mining module	D3.4	Exploitation by BM-Y!
8	Topic-based classification module	D4.3	Freely available
9	Semantic content integration framework	D4.4	Open source (OS)
10	Multimodal indexing and retrieval module	D4.3	Freely available
11	Semantic representation Infrastructure	D5.4	Open source (OS)
12	Hybrid reasoning module	D5.4	Exploitation by ONTO
13	Decision Support module	D5.4	Exploitation by ONTO
14	Summarisation module	D6.3	Open source (OS)
15	Final system	D7.7	Exploitation by the consortium

The project outcomes have been updated according to D7.1 Roadmap deliverable as follows:

#	Module	Outcome
1	Content Extraction Module (WP2)	Named entities extraction component
2		Dependency parsing component
3		Concept extraction component
4		Automatic speech recognition component
5		Multimedia concept and event detection component
6		Machine translation component
7	User and context-centric content analysis module (WP3)	Context extraction and representation component
8		Polarity and sentiment extraction
9		Information propagation and social interaction analysis
10	Multidimensional content integration (WP4)	Multimodal indexing and retrieval component
11		Topic-based modelling component
12		Mapping discovery and validation component
13		Content alignment and integration component
14	Semantic reasoning and decision support (WP5)	Data infrastructure module
15		Semantic representation infrastructure management system module
16		Decision support system module
17	Content summarisation and delivery (WP6)	Extractive summarisation
18		Abstractive summarisation
19	System development and integration (WP7)	Data crawling module
20	Final System	MULTISENSOR

In the following we provide a detailed description and the exploitation perspective for each of these modules, which are considered as the main outcomes and the foreground of the project.

3.1 Content Extraction Module (WP2)

3.1.1 Named Entities Extraction Component

Module Description	The Named Entities recognition Component identifies names in texts. Names are words which identify objects, like 'Maastricht Treaty', 'Berlin', 'Siemens'. Names belong to different types. The component can identify the following entity types: • persons • locations • organisations, divided into companies and institutions • amounts • dates It will be available for 5 languages (English, French, German, Spanish and Bulgarian).
Innovation Description	Unlike other NER components which build on only shallow analysis techniques, the approach in MULTISENSOR is to choose a technology which can be extended. This approach allows for deep analysis of texts, which is expected to result in higher precision of results and easier adaptability towards new domains.
IP rights	The NER component consists of three software components: Sentence splitting, tokenization, and NE recognition. The NE recognition uses three components: local parser, text analyser (for co reference determination etc.), and output generation. IP rights for the software components are with Linguatec. The lexica and gazetteers belong to the respective partners in the project who created them.
Foreseen license	Commercial licence
Alternative solution	• Stanford NER: Supports English, German, Spanish, Chinese. License only for non-commercial applications. • GATE Information Extraction Component ANNIE, with grammar JAPE. Grammar language supports only adjacent constituents, no longer-term dependencies. • OpeNER: Origin is an EU Research project, consisting of many components with unclear license conditions. Support status is also unclear. • Open Calais: a service offered by Thomson Reuters. Currently not able to support development of new domains or languages.
Adaptability and extensibility	Because of the modular NER architecture it can be adapted to other domains and languages with only limited effort. The software components are basically language-independent, so that only the resources need to be adapted, i.e. NER lexicon (annotated gazetteer) and grammar rules. The foreseen effort for

	adaptation can be evaluated when the component with its required resources is completed. During the project the component runs in the Linguatec cloud as this allows for easy access and improvement.

3.1.2 Dependency Parsing Component

Module Description	The statistical parsing module takes a natural language sentence and outputs either its surface- or deep-syntactic dependency structure (depending on the choice of the user). The module shows high quality performance for a number of languages – in particular English and Spanish. Its accuracy decisively depends on the size of the training corpus and the quality of its annotation.
Innovation Description	The deep parsing component of the module is a unique service to the market given the fact that no comparable parsers are available. We expect it to be of high interest to several downstream applications such as deep machine translation, information extraction, paraphrasing, etc.
IP rights	UPF is the owner of the deep parsing component. The IP right of the surface parsing component (which is used in a pipeline with the deep parser) belongs to a third party.
Foreseen license	GNU GPL v3[16]
Alternative solution	–
Adaptability and extensibility	The dependency parsing module itself is language-independent. To apply it to other languages than those in MULTISENSOR, text corpora in the corresponding languages must be annotated with surface- and deep-syntactic structures. If a surface-treebank is already available, a semi-automatic mapping of the surface structures to deep-syntactic structures can be performed. The effort for the annotation may range thus from 3 PMs to 12 PMs.

3.1.3 Concept Extraction Component

Module Description	The concept extraction module operates on plain text, drawing upon a number of standard semantic resources such as Babelnet, Babelfy, Framenet, etc. It outputs concepts of the analysed text and relations between them. In its basic variant, it consists of off-the-shelf components, including a surface dependency parser and the Semafor tool. In its advanced variant, it will consist of UPF's deep parser and an own Semafor-like component.
Innovation Description	The innovation of the advanced version of the module is in the combination of deep parsing technologies with semantic processing. It can be considered a novel service for the ICT market.
IP rights	UPF is the owner of the IP rights of the advanced variant of the concept extraction module. The IPR rights of the components of the basic variant are with third parties.

[16] http://www.gnu.org/licenses/gpl.html

Foreseen license	GNU GPL v3[17]
Alternative solution	
Adaptability and extensibility	In order to adapt the concept extraction module to new domains, sufficient language-specific training material is required. The annotation effort depends on the language and linguistic difference of the new domain with the MULTISENSOR domains. An effort of 3 to 6 PMs is realistic.

3.1.4 Automatic Speech Recognition Component

Module Description	Automatic speech recognition (ASR) is employed within the MULTISENSOR project to provide a channel for analysis of spoken language in audio and video files. It transforms speech signals into a sequence of phonemes and words. The recognition quality depends on different factors such as speaker and channel variability, background noises, audio frequency spectrum, quality of microphone, or difficulty in differentiation between speech and non-speech events. The module will be available for 5 languages (English, French, German, Spanish and Bulgarian).
Innovation Description	The ASR technology used in MULTISENSOR is speaker-independent and uses an open-vocabulary approach (recognition of unknown words based on sub-word units). It employs a series of state-of-the-art techniques: continuous density HMMs for the acoustic modelling; MFCC or PLP feature extraction with support of LDA and VTLN; speaker adaptation by the means of CMLLR; time-synchronous left-to-right beam search strategy for the decoding. The advanced versions will be adapted by using in-domain data, as much as the project partners manage to collect. Another innovation is the integration of the results from the named entities recogniser. This is expected to result in better recognition of proper names in spoken language.
IP rights	IP rights for the software components and resources are with Linguatec. Some of the background components are with RWTH Aachen.
Foreseen license	Commercial licence
Alternative solution	A commercial ASR solution is available from Nuance Communications. As for Open Source alternatives, there are ASR engines such as SPHINX (from Carnegie Mellon University, USA) and JULIUS (from Nagoya Institute of Technology, Japan). A more recent open source engine is KALDI. However, ASR engines still require access to audio and text data and the generation of domain specific language and acoustic models.
Adaptability and extensibility	Adaptation to other domains requires large in-domain data to train appropriate Language Models. Adaptation to other languages is substantially more complex as it involves extensive recordings of native speakers together with precise transcriptions.

[17] http://www.gnu.org/licenses/gpl.html

3.1.5 Multimedia Concept and Event Detection Component

Module Description	**Functionality, Input/Output:** The multimedia concept and event detection component receives as input a multimedia file (i.e. image or video) and computes degrees of confidence for a predefined set of concepts and events. In order to achieve this, the component incorporates various procedures, such as video decoding (applicable for video files only), feature extraction and supervised classification. **Dependencies**: OpenCV and vlfeat libraries, ffmpeg, ffprobe
Innovation Description	The main innovation of the component is the utilization of state-of-the-art techniques in all the above mentioned incorporated procedures (video decoding, feature extraction, classification) for automatically annotating a multimedia file based solely on visual content. This module could be exploited either as service or standalone and integrated in a media monitoring and multimedia management product.
IP rights	CERTH is the owner of the innovation
Foreseen license	Open Source (GNU General Public License, version 3 (GPL-3.0))
Alternative solution	N/A
Adaptability and extensibility	The component is language independent and given a reasonable amount of time (approximately 3 PMs/10 concepts), it could be trained and extended/adapted to other concepts and events. In order to extend it to additional concepts the module requires annotated data (i.e. images and videos that depict this concept/event and training of the predictive models).

3.1.6 Machine Translation Component

Module Description	Automatic machine translation (MT) is employed within MULTISENSOR to provide the translation of the summarisation results in the end of the content analysis and summarisation chain and to enable full-text translation on-demand. In the first case, the translations will be produced at the end of the analysis/summarisation process and will be stored together with the summaries. In the second case, the translations produced by MT provide the input for the text analysis chain and follow the same analysis procedure as the input from original text sources in the required language. The languages covered by MT in the MULTISENSOR project will be English, French, German, Spanish and Bulgarian.
Innovation Description	In MULTISENSOR a Statistical Machine Translation (SMT) approach and the machine learning techniques associated to it are applied. A phrase-based translation model is used, i.e. instead of learning the translation word by word, larger word sequences (up to 7 words) are being taken into account. Thus larger contexts, different word orders in source and target, as well as distant dependencies are taken into account. Another innovation is the integration of the results from the named entities recognizer. This is expected to result in

	better translation of proper names.
IP rights	The MOSES translation decider is open source. The IP rights for the generated phrase tables and language models are with Linguatec.
Foreseen license	MOSES decoder: open source (LGPL license) Language Models, phrase tables and lexica: commercial license from Linguatec
Alternative solution	As an alternative to statistical MT there are several rule-based MT systems (e.g. Systran, Promt, Linguatec Personal Translator), all of which need a commercial license. With OpenLogos there is also an open source alternative for rule-based MT. However it currently only supports English and German as source languages.
Adaptability and extensibility	Adaptation to other domains requires bilingual in-domain texts to train appropriate phrase tables. Adaptation to other languages requires in addition substantially larger bilingual training texts. By using English as a pivot language, however, it will be possible to support 12 additional translation directions (e.g. French-German, Bulgarian-Spanish).

3.2 User and Context-Centric Content Analysis (WP3)

3.2.1 Context Extraction and Representation Component

Module Description	The Context analysis module is designed to extract or derive the set of contextual features associated with a media item. Given a media item, we consider the following contextual features: • Author: an entity responsible for the creation of the item content • Source: an entity responsible for making the item available • Title: a name given to the media item • Keywords: a set of phrases describing the topic of the item • Genre: the style or type of the item • Category: a classification of the item according to its content • Date: a date associated with the creation or availability of the item • Location: a location indicating where the item was created • Literary style: a metric of language formality • Language: the language of the content of item
Innovation Description	The module can be used by a company to extract contextual features appropriate for describing an item such as a news article or a blog. The innovation of this component will lie in the potential use of analytics based on the contextual features.
IP rights	The owner of the innovation is BM-Yahoo!
Foreseen license	The software supporting this module will become available under the MIT free software licence on github.
Alternative solution	N/A

Adaptability and extensibility	For the features extracted from the metadata, the module can be considered language-independent. For the features extracted from the text, we assume that the text would be in English (either originally or translated by the relevant component).

3.2.2 Polarity and Sentiment Extraction

Module Description	The Polarity and Sentiment Extraction module is designed to perform efficient and effective sentiment classification. More specifically, it involves two main subcomponents: a) sentimentality (also known as subjectivity) detection - a text segment is classified as either subjective or objective b) polarity detection - a text segment is classified as either having positive or negative sentiments
Innovation Description	The proposed module will offer a tailor-built, domain-specific sentiment analysis solution, trained on specific domains of interest and value to Yahoo, but also extensible to new ones. To this end, we will employ a machine-learning approach using a wide range of NLP features, so that the classification of news content containing opinions is done with simple yet effective operations.
IP rights	The owner of the innovation is BM-Y!
Foreseen license	The technology will be made available "for non-commercial purposes" or "for research purposes". The consortium members will have access to the source code and the final module, to be able to use the technology to fulfil the requirements of the project.
Alternative solution	SentiStrength[18], SentiWordNet[19]
Adaptability and extensibility	Contrary to lexicon-based, domain-independent solutions, which are not straight-forward in how they can be effectively extended to other domains, the proposed module will be a tailor-built, domain-specific, machine-learning solution trained on specific domains of interest. In principle, the Polarity and Sentiment Extraction module is language-dependent and should be easily extensible to new domains, given the availability of annotated text corpora with sentiment features. However, certain techniques and features (e.g., syntactic features) may introduce language dependencies.

3.2.3 Social Media Mining Module

Module Description	The Social Media Mining module is designed to perform analysis on users with respect to their potential influence on online audiences, and in particular, on Twitter. The ContributorAnalysis module can be summarized as follows:

[18] http://sentistrength.wlv.ac.uk

[19] http://sentiwordnet.isti.cnr.it

	• Input: a string representing the Twitter screen-name of the user of interest (e.g., @barackobama) • Output: a list of attributes associated with the particular user (description, location, language, interests, number of posts, friends, followers, network and retweet influence scores) • Functionalities: Besides crawling Twitter for the basic profile data of the user, the module computes two influence scores: 1. network influence score based on number of followers and followees 2. retweet influence score based on retweet counts Finally, it also provides a summary of the interests of the user, by analysing his/her activities on the social network. • Dependencies: User authentication Twitter API keys and twitter4j library.
Innovation Description	The module can be used to retrieve information publicly available on Twitter users. It provides multiple authority scores computed by leveraging different signals (number of followers, followees, and retweet counts). Similar scores are also available by social media analytics companies such as Klout; however: a) scores on other platforms are available only for some and not all users b) scores cannot be directly exported in other systems or knowledge bases c) the exact calculation algorithm of the score is not disclosed, while our computation is transparent.
IP rights	The owner of the innovation is BM-Yahoo!
Foreseen license	The software supporting this module is available under the MIT free software licence on github.
Alternative solution	N/A
Adaptability and extensibility	The module is domain and language independent and can be expanded in several ways: 1. Integration of additional signals in the computation of authority score 2. Refinement of the interest analysis

3.3 Multidimensional Content Integration (WP4)

3.3.1 Multimodal Indexing and Retrieval Component

Module Description	Functionality, Input / output: The multimodal indexing and retrieval module involves the development of a multimedia data representation framework that allows for the efficient storage and retrieval of socially connected multimedia objects. Specifically, it develops a representation model (called SIMMO) for holding several dimensions of the multimedia information. Moreover, it contains an indexing structure for holding and retrieving efficiently the multimodal entities of the multimedia information. Dependencies: MongoDB, Morphia framework

Innovation Description	The main innovation of the component is the development of the indexing and retrieval module based on SIMMO (Socially Interconnected MultiMedia-enriched Objects) model, which has the ability to fully capture all the content information of interconnected multimedia objects, while at the same time avoids the complexity of previously proposed models. This product will be used for media monitoring purposes.
IP rights	CERTH is the owner of the innovation
Foreseen license	Open Source (GNU General Public License, version 3 (GPL-3.0))
Alternative solution	N/A
Adaptability and extensibility	The component is language and domain independent. If the data representation requirements that stem from the characteristics of the SIMMO model are taken into account and met, then it is plausible to extend the component to include additional fields of information with relative ease (e.g. 1-2 PMs).

3.3.2 Topic-based Modelling Component

Module Description	Functionality, Input/Output: This component includes two basic functionalities: a) category-based classification and b) topic-event detection. The component receives as input multimodal features and provides as output a) the degree of confidence of a number of categories for a specific News Item (for category-based classification) b) a grouping for a list of News Items based on the existence or not of a number of topics / events (for topic-event detection). Dependencies: R (statistical programming language)
Innovation Description	The use of novel late classifier fusion approaches (for the category-based classification task) and the utilisation of multimodal clustering techniques (for the topic-event detection task). This module could be exploited either as service or standalone and integrated in a media monitoring and multimedia management in order to provide automatic category tagging (classification) and grouping of heterogeneous multimedia (clustering).
IP rights	CERTH is the owner of the innovation
Foreseen license	Open Source (GNU General Public License, version 3 (GPL-3.0))
Alternative solution	N/A
Adaptability and extensibility	The component, if trained accordingly, can receive as input features from any number of modalities. The extension/adaptation procedure of the component to other domains is considered relatively easy (About 2 PMs for any new domain set (regardless of the domain number)). It must also be taken into account that the component receives its input from the indexing structure of the multimodal indexing and retrieval component, therefore any extension/adaptation to other domains should also consider the indexing component. The approach is language agnostic and it assumes that each

	multimedia document is represented with vectors (e.g. bag of words).

3.3.3 Mapping Discovery and Validation Component

Module Description	**Functionality, Input/Output**: The mapping discovery and validation component deals with discovering and registering in an automated way, valid mappings between the concepts and properties of two ontologies. The ontologies of WP5 are considered for mapping, since a manual discovery of the mappings between the ontologies is a tedious process, especially if the latter are big. The component uses string, lexical and structural similarities combined in a late fusion approach in order to estimate the similarity between two concepts of the ontologies. The mappings are further semantically validated for consistency.
Innovation Description	The main innovation of this component lies in the development of a weighted late fusion approach for combining the different similarities for the metrics that are employed. This application can be used as administrative support for ontology engineers and ICT companies that deal with solutions based on semantics.
IP rights	CERTH is the owner of the innovation
Foreseen license	Open Source (GNU General Public License, version 3 (GPL-3.0))
Alternative solution	N/A
Adaptability and extensibility	The component can be extended by integrating new similarity algorithms that take advantage of different metrics. For instance in order to include a new existing matching algorithm a work of 2-3 PMs would be required. This application is domain and language independent.

3.3.4 Content Alignment and Integration Component

Module Description	Functionality, Input/Output: The content alignment and integration component deals with the discovery of relations or inconsistencies between content items in the knowledge base. Identifying hidden relations in content enriches the knowledge base which in turn enables enriched results. On the other hand, identifying inconsistencies between content allows identification of noisy entries in the knowledge base, e.g. false assertions, which should be further considered. Early versions of the module use querying strategies for content alignment, while more advanced methods that employ rule-based approaches are being developed. Dependencies: GraphDB or any semantic repository
Innovation Description	The main innovation of this component lies in the development of a content-oriented approach to identify hidden relations or inconsistencies between content items. The approach is generic so that it can be adapted to a variety of domains. It can be used as part of a media monitoring product.
IP rights	CERTH is the owner of the innovation
Foreseen license	Open Source (GNU General Public License, version 3 (GPL-3.0))
Alternative	N/A

solution	
Adaptability and extensibility	The component can be adapted to run on any semantic repository, while it can be tuned to specific domains of interest. In order to adapt it for a different domain an effort of 2-3 PMs is expected. The component is language independent.

3.4 Semantic Reasoning and Decision Support (WP5)

3.4.1 Data Infrastructure Module

Description	Ontotext provides GraphDB-Enterprise as a semantic data infrastructure layer for the purposes of MULTISENSOR. It is a high performance system implemented in Java, which support storing, querying and processing structured data formatted according to the RDF standards and is packaged as Storage and Inference Layer (SAIL) for the Sesame framework. It is based on the Ontotext's TRREE – native RDF rule-entailment engine. GraphDB is world leader among the structured data repositories in terms of volume of data and loading/inference speed. One of the main advantages is the in-memory reasoning implementation- the content of the repository is loaded and maintained in the main memory, which allow for efficient retrieval and query answering.
Innovation Description	Ontotext will implement and deliver four different types of reasoning, which currently are not supported by GraphDB. There is no product on the market which support these inference techniques together. • Parallel inference • Hybrid reasoning • SPARQL-MM • GeoSPARQL
IP rights	Owner of this innovations is Ontotext.
Foreseen license	GraphDB is available under an RDBMS-like commercial license on a per-server-CPU basis.
Alternative solution	As alternative users can use Sesame, which is an open source framework for storing, querying and analysing RDF. It is implemented in Java by Aduna and is available under the GNU Lesser GPL license[20].
Adaptability and extensibility	GraphDB can be adapted to work on the cloud very easily. ONTO already has such kind of project named S4, where we offer Data-as-a-Service. GraphDB has the ability to work with many different languages, so all the innovations support these languages tool.

3.4.2 Semantic Representation Infrastructure Management System Module

Description	GraphDB Workbench is provided as Semantic representation infrastructure management system. It is our recommended web-based administration tool. The user interface is similar to the Sesame Workbench web app, but provide

[20] http://www.gnu.org/copyleft/lesser.html

	more functionalities, better user experience and clean design. Some of the additional features are: • Query monitoring with the possibility to kill a long running query. • Better SPARQL editor based on YASGUI[21] • Connectors administration presented only in the enterprise edition. GraphDB Workbench can be used as a SPARQL endpoint and as an administration tool for managing repositories, executing queries and updates. It contains user management module which allow to set different kind of restrictions over the repositories like read/write for different user groups. In addition to the SPARQL queries, you can perform Full Text Search over the data, but with arrangement that such index exists.
Innovation Description	The main innovation here is the new improved UI. This new UI will give us great new functionalities and management capabilities compared with Sesame Workbench.
IP rights	Owner of this innovations is Ontotext.
Foreseen license	GraphDB Workbench is part of the GraphDB distribution so is available under an RDBMS-like commercial license on a per-server-CPU basis.
Alternative solution	As alternative users can use Sesame Workbench, which provides all basic functionalities for managing, querying and updating the semantic repository. It is open source and is available under the GNU Lesser GPL.
Adaptability and extensibility	Currently the GraphDB Workbench is available only in English language. GraphDB Workbench can be easily extended depending on the specific use case.

3.4.3 Decision Support System Module

Description	ONTO is going to implement restful web services to expose the functionalities of the decision support system which is built on the top of the semantic repository. This first version of the system will be very basic and will provide limited capabilities to support the third use case – Internationalization. In the process of development we identified the most important statistical indicators from The World Bank and Eurostat which are well described in D3.2. Based on the data in this indicators and datasets like DBpedia and Geonames we can develop specific SPARQL query templates. We also added support for Google charts on the MULTISENSOR SPARQL endpoint[22], so the information of the queries can be very easily visualized. Another addition to the whole system but especially to the decision support is the implementation of GeoSPARQL standard which will help the users to work with geospatial objects. These template queries together with the restful web services, the GeoSPARQL support and the Google chars are in the core of our first version of the decision support system.
Innovation Description	• Develop SPARQL query templates which will help to get specific results depending on the decision support use cases

[21] http://laurensrietveld.nl/yasgui

[22] http://multisensor.ontotext.com/sparql

	• Develop new restful web service, to retrieve the results generated by the SPARQL templates. The queries are identified by ID • Develop new UI functionalities to support Google charts, so the results of the queries can be easily visualized, which will help in the process of decision making.
IP rights	Owner of these innovations is Ontotext.
Foreseen license	As part of the GraphDB Workbench it will be available under an RDBMS-like commercial license on a per-server-CPU basis.
Alternative solution	As an alternative, one can use Sesame Workbench or Yasgui[23], but neither of these alternatives has the needed functionalities and the capabilities ONTO is going provide.
Adaptability and extensibility	Currently GraphDB Workbench which takes the role of management application is available only in English language. GraphDB Workbench can be easily extended depending on the specific use case, but precise estimation can't be given because it highly depend by the specific use case.

3.5 Content Summarisation and Delivery (WP6)

3.5.1 Extractive Summarisation

Module Description	The module operates on plain text or collection of texts, assessing the relevance of individual sentences to the summary accordance to a series of quantitative metrics. The most relevant sentences are selected and delivered to the user. The core of the module is the SUMMA summarization toolkit.
Innovation Description	The innovation of this module consists, first of all, in the metrics developed in the context of MULTISENSOR. From the perspective of the market, this module provides tuning facilities to an existing service.
IP rights	IP rights of the SUMMA summarization toolkit belong to a third party (Dr. Horacio Saggion)
Foreseen license	Proprietary license
Alternative solution	The abstractive summarisation component in section 3.5.2.
Adaptability and extensibility	To adapt the extractive summarization module to new domains, training material consisting of a sufficient number of sample summaries (along with the original texts) of high quality is needed. No estimation of the effort of the compilation of such summaries can be given, since it depends on existence of quality annotated summaries for specific domains and languages. The effort for retraining of the summarization module is minimal.

[23] http://yasgui.org/

3.5.2 Abstractive Summarisation

Module Description	The module operates on the ontological representations of the content distilled from a given text and outputs a summary of this text in one of the languages of MULTISENSOR. It consists of three main components: (1) content selection, (2) discourse structuring, and (3) text generation. For sentence generation within the third component, a rule-based, statistical or hybrid generator can be chosen.
Innovation Description	The module possesses an innovative architecture (in that it is, de facto, a genuine content-to-text generator) and innovative realizations of the individual components. The module is thus a novel service for the ICT market.
IP rights	UPF possesses the IP rights of the abstractive summarization module.
Foreseen license	NU GPL v3
Alternative solution	N/A
Adaptability and extensibility	The content selection component of the module can be operated on any RDF/OWLIM content structures without the need of adaptation. The discourse structure module will require some adaptation to new domains; it is, however, language-independent. The adaptation of the rule-based text/sentence generator presupposes the compilation of language-specific grammatical and lexical resources at different levels of the linguistic description. The size (and thus coverage) of these resources depends on the nature and verbosity of the targeted summaries. The adaptation of the stochastic sentence generator requires the annotation of text corpora with linguistic structures at different levels of abstraction (surface-syntactic, deep-syntactic, and semantic) for each language that is to be covered by the summarizer. For an acceptable performance of the stochastic sentence generator, training treebanks with between 3,500 sentences (with a very high quality of annotation) and 10,000 sentences are needed. The estimated cost is about 1PM per 1000 sentences of high quality annotation.

3.6 System development and integration (WP7)

3.6.1 Crawlers and data channels infrastructure

Module Description	The Crawler component is the process responsible of collecting source material for use by the MULTISENSOR platform. It runs regularly on a set schedule, going over a set of manually selected content sources, and sends the retrieved material through the analysis pipeline for extraction of intelligence. Sources includes print and online news sites, social media, and audio and video sources (online radio, podcasts, tube sites, etc.). Raw data is normalised into common JSON format and stored in the appropriate repository. The crawler depends on two sub-components or “Collectors”, the Media Collector and the Site Collector.
Innovation Description	The crawler infrastructure is an innovative product that may be used as a framework for implementing other “Collectors”, providing a convenient system for scheduling the crawling process and converting the harvested content to a

	JSON output. The crawling tasks are performed using standard techniques and off-the-self tools. In that sense there is little room for innovation.
IP rights	everis, PR and BM-Y!
Foreseen license	Open Source (BSD) for the Crawler infrastructure (everis) and Media Collector (BM) and proprietary rights for the Site Collector (PR).
Alternative solution	With **Nutch**[24] BM-Y! has setup an open source crawler. Therefore, there are no proprietary rights involved. Other well-known open-source crawling applications are: **Scrappy**[25], and **Heritrix**[26]. Media currently delivered through PRs proprietary crawlers could be substituted through such an open source crawling solution.
Adaptability and extensibility	This component may be adapted to other "Collector" sub-components. If these respect the interface of the Crawler, integration should be perfectly feasible.

3.7 Final System (WP7)

Description	The final system includes the MULTISENSOR platform integrating the aforementioned components and modules. Specifically the final system includes the 3 UC Portals, integration services, business services, and associated repositories. The designed architecture follows the latest trends and developments in this area, by proposing a decoupled, layered, service-oriented system, and the use of lightweight RESTful API for accessing the different services. The central infrastructure has also been designed to run in a cloud environment. The system makes use of JSON objects, RDF data, RESTful APIs and repositories such as Elastic Search, MongoDB and GraphDB has become an excellent opportunity to develop state-of-the-art solutions for integrating these elements.
Innovation Description	The final system includes 3 innovative products: a) portal for journalism, b) a portal for commercial media monitoring and c) a portal for SME internationalisation decision support. According to the market analysis performed in D8.2 and in chapter 2 it is clear that the 3 products have a very high innovation potential. With respect to a) and b) the MUTLISENSOR platform is an innovative product that allows for multilingual and multimedia search across several different sources, including both quick an in-depth analysis support through summarisation, sentiment analysis, network analysis and concept detection could work, just as there doesn't seem to be anything as comprehensive in one single interface. With respect to c), MULTISENSOR platform comprises an innovative product that can provide decision support in order to assess potential international investments by considering internationalisation indicators and by providing unified access to multilingual

[24] http://wiki.apache.org/nutch/FrontPage

[25] http://scrapy.org/

[26] https://webarchive.jira.com/wiki/display/Heritrix

	content.
IP rights	All partners
Foreseen license	The portals are open Source (BSD), however each component is governed by each specific license as described above.
Alternative solution	As the system has been designed and developed originally following very specific requirements, no substitute products exist to our knowledge.
Adaptability and extensibility	The SOA approach that has been followed allows for further functional enhancements of the system, provided the interface standards defined in the project are respected. That is, new services and functionalities should adapt to the system, not the other way. With regard to adaptability to new domains or languages, the system depends on all the aforementioned modules for which specific details have been provided in the previous tables with respect to their extensibility and adaptability. The fact that the infrastructure is cloud-based also makes scalability possible without major challenges. On the other hand, it is also possible to implement the system in a local infrastructure. This approach will not limit the system's functionalities in any way, provided the infrastructure is properly dimensioned. However, the performance may be slower due to network capacity. During the project, a cloud infrastructure approach is adopted. The main reason for this is that a flexible solution is needed to accommodate the system's evolving requirements. At the beginning of the project, it was rather difficult to foresee the required capacity, and therefore a flexible solution was essential. A cloud infrastructure can be scaled up and down easily and almost instantly. This has the benefit of aligning infrastructure and its associated costs with its usage. Moreover, everis does not have data centres that could fulfil the project's requirements. It is beyond the scope of everis business to own such infrastructures, and these are always subcontracted from third parties. All the above also applies to the commercialisation phase of MULTISENSOR. The user requirements and the load these may put into the system are unknown, and therefore a scalable solution is still required. A cloud infrastructure may be optimised, for example, to a situation where the system is put on "stand-by" i.e. no crawling nor processing takes place. In this scenario, the system's capacity needs would be much lower than during the duration of the project, and a cloud architecture will ensure that the infrastructure is not over-dimensioned. It is important to note that some of the services – e.g. machine translation and GraphDB database – reside in the partner's premises and therefore outside the main platform. However, the maintenance of each service/module has been discussed above in detail and in the case of commercial products alternative open source solutions have been suggested.

4 CONCLUSION

In an attempt to increase the chances of the exploitation of the project results after the project's end, we have first tried to map the specific challenges which journalists, media monitoring companies and SMEs on the brink of internationalization are commonly facing. These are represented in the 3 MULTISENSOR use cases. We have then laid the exploitable foreground by analyzing the current market situation and what kind of existing tools are already available with the goal of identifying potential market sectors and niches for the project results.

Our extended market analysis has shown a broad spectrum of available comparable tools already covering a wide range of MULTISENSOR functionalities. Yet, there doesn't seem to be anything on the market that offers MULTISENSORs wide range of functionalities such as multilingual search, in-depth analysis support through summarisation, sentiment analysis, network analysis and concept detection in one single comprehensive interface. Especially our analysis in regards to the SME Internationalization use case has uncovered a unique selling proposition and strong differentiation points versus available solutions on the market: a clear focus on providing actionable business intelligence for export-oriented SMEs.

Finally we have presented a clear overview (including descriptions, licences, etc.) of the technological modules and services that are developed in the project in order to facilitate uptake by interested parties.

During the next year the consortium will carefully review each of the options identified and propose the final exploitation strategy (Del. 9.7) based on the results of this analysis.