A CIP-PSP funded pilot action
Grant agreement n°325188

Deliverable	**D6.2.2 – ACDC Social Analytics tools, Online tools**
Work package	WP6
Due date	31/07/2015 (resubmission)
Submission date	30/11/2015 (resubmission after final review)
Revision	0.10
Status of revision	Final
Responsible partner	Engineering Ingegneria Informatica
Contributors	Angelo Marguglio (EII), Giuseppe Vella (EII), Paolo Roccetti (EII), Véronique Pevtschin (EII), Ioana Cotoi (EII)
Project Number	CIP-ICT PSP-2012-6 / 325188
Project Acronym	ACDC
Project Title	Advanced Cyber Defence Centre
Start Date of Project	01/02/2013

Dissemination Level	
PU: Public	✓
PP: Restricted to other programme participants (including the Commission)	
RE: Restricted to a group specified by the consortium (including the Commission)	
CO: Confidential, only for members of the consortium (including the Commission)	

Version history

Rev.	Date	Author	Notes
0.1	27/01/2014	EII	Creation of first draft
0.2	28/01/2014	EII	Refined ToC
0.3	30/01/2014	EII	Contribution to section 4
0.4	31/01/2014	EII	Contribution to section 3, updated section 4
0.5	5/2/2014	EII	Review / addition of executive summary, introduction and conclusion
0.6	02/02/2015	EII	Updated section 5 to reflect actual implementation of ACDC Analytics. Updated sections 1, 3 and 6
0.7	11/02/2015	EII	Final changes to the document
0.8	20/07/2015	EII	Updated section 5 to the new implementation of ACDC Social Analytics
0.9	18/11/2015	EII	Updated section 5 with actual analytics and results regarding the use of the portal
10	21/11/2015	EII	Updated executive summary, introduction and conclusions adding next steps

Table of contents

1. Executive summary 6
2. Overview of the link between WP5 and WP6 deliverables (M24 deadline) 8
3. Introduction 10
4. Social Analytics Overview 11
   4.1. Social Analytics purposes 11
   4.2. Kind of Analyses & related inputs 12
   4.3. Output Representations 15
   4.4. Rise of social analytics tools 18
5. Modelling ACDC Social Analytics 25
   5.1. Purposes of the ACDC Analytics 25
   5.2. ACDC Analytic dashboard design 26
      5.2.1. User contents 26
      5.2.2. Data Sharing 32
   5.3. ACDC Social dashboard implementation 36
   5.4. Monitoring the ACDC User Community 37
      5.4.1. Analysing the ACDC Social Interaction 37
      5.4.2. Analysing the ACDC Data Sharing 42
      5.4.3. Analysing the ACDC Botnet Metrics 43
6. Conclusions 46
7. References 48

Table of figures

Figure 1 - Example of Timeline graph 15
Figure 2 - Example of GEO Map 16
Figure 3 - Example of Table 16
Figure 4 - Example of Pie chart 17
Figure 5 - Example of Bar chart 17
Figure 6 - Example of Heat map 18
Figure 7 – Stakeholder activities, overall view 28
Figure 8 – Stakeholder activities, detailed view 28
Figure 9 – Resource activities, overall view 30
Figure 10 - Activities by resource type over time 31
Figure 11 - resource activities, detailed view 31
Figure 12 - DAM Usage Totals 33
Figure 13 - DAM Read and Write keys 33
Figure 14 - DAM Stakeholders by group 34
Figure 15 - DAM API Keys by group 35
Figure 16 - DAM sharing policies by group (sample) 36
Figure 17 – Results regarding the Reports 37
Figure 18 – Results regarding the Stakeholders activities 38
Figure 19 – Stakeholder Analytics ATOS 39
Figure 20 – Stakeholder Analytics ENG 39
Figure 21 – Comparison on the stakeholders results 39
Figure 22 – Results on «Documents» 40
Figure 23 – Results on «News» 40
Figure 24 – Results on «Forum» 41
Figure 25 – Results on «Tools&Services» 41

Figure 26 - DAM Statistics overview 42
Figure 27 - DAM Sharing Policies distribution 43
Figure 28 - Gaps in time graph 44
Figure 29 - Unique IP per Country example 45

Table of tables

Table 1 – Overview of the WP5 – WP6 deliverables over the 30 months of operation 9
Table 2 - Capabilities of Profiled Social Analytics Vendors in Relation to Less Commonly Examined Information Sources 20

Glossary

ACDC	Advanced Cyber Defence Centre
BI	Business Intelligence
CERT	Computer Emergency Response Team
CSIRT	Computer Security Incident Response Team
DoW	Description of Work
EC3	European Cybercrime Centre
ECI	European Critical Infrastructure
EEC	European Economic Community
ENISA	European Network and Information Security Agency
EU	European Union
FS-ISAC	Financial Services - Information Sharing and Analysis Centre
ICANN	Internet Corporation for Assigned Names and Numbers
ICT	Information and Communication Technologies
IETF	Internet Engineering Task Force
IGF	Internet Governance Forum
INTUG	International Telecommunications Users Group
ISAC	Information Sharing and Analysis Centre
ISACA	Information Systems Audit and Control Association
ISIC	International Standard Industrial Classification of All Economic Activities
ISP	Internet Service Provider
LAP	London Action Plan
LEA	Law Enforcement Authority
LoI	Letter of interest
MAAWG	Messaging Anti-Abuse Working Group
NACE	Statistical Classification of Economic Activities in the European Community
NSC	National Support Centres (ACDC relay nodes)
RIPE (NCC)	RIPE Network Coordination Centre
SCADA	Supervisory Control and Data Acquisition
US	Unites States

1. Executive summary

One of the major contributions to the out-reach approach of the ACDC project in fighting botnets across Europe was the creation of a stakeholders' community. In order to support the sharing of information and knowledge, improve communication between stakeholders all over the Europe, share data and participate to experiments, ACDC deployed a community platform, used as the main channel for community members to interact and to take part in concrete activities.

To better support these interactions, ACDC introduced social analytics as a set of tools to ease the analysis of the ACDC community in terms of growth, evolution and behaviour. Therefore, ACDC Analytics provide figures about the community participants, the contents actually stored in the community portal, the social interaction, the data interaction using the CP / the CCH, statistics about the data in the CCH etc.

In order to specialise analytics to the ACDC context, this deliverable introduces, in section 4, an overview of social analytics in terms of purposes and visualisations options, as well as the main tools available today from larger IT providers.

In section 5, the document details the specific reports implemented for the ACDC context, focusing on the purposes of the ACDC Analytics, the dashboard design and the implementation of it. Moreover, in the subsection "Monitoring the ACDC User Community" the document reports on the actual analytics about portal usage and the visualization of these results.

ACDC is a community focusing on fighting botnets, and composed by organisations with very different profiles and involvement levels. Due to this, instead of focusing only on contributions of single companies or individuals, ACDC Analytics also reports on the status of the ACDC community, as a whole. Therefore, the ACDC Analytics allows managers to explore the community status from different points of view: user contents, data sharing and botnet metrics. Each topic has a set of associated reports that can also be compared to derive further insights about the status of the community.

In particular, the user content section of the application allows exploring and visualizing the actions on the contents of the portal. The analysis can follow two approaches: the *Stakeholder approach* and the *Resource approach*. In the *Stakeholder approach*, at top level, the social interaction is shown grouped by stakeholders (i.e. organizations), and detailed at the level of a single user. In the *Resource approach*, the social interaction is shown grouped by content types (i.e. news, initiatives, messages in forum, etc.) at top level, and detailed at the level of a single content. This enables different kinds of evaluations by community managers, for instance identify lack of involvement of a stakeholder, or assess which resources generated more interaction (thus more interest) in the community.

Besides the user content analysis, the application also contains statistics on the data sharing related to the adoption and usage of the Data Access Manager, the tool of the community portal that enables the ACDC community to access the Central Clearing House (CCH) component to retrieve and provide data about cyber-attacks. The data sharing section is useful to assess the participation of the ACDC community to the data exchange that is at the base of the ACDC approach to fight botnets. This is made possible by a set of graphs partitioning the DAM usage by stakeholder type, by the number of access keys that are used to access the CCH and by sharing

policies that each stakeholder defined to control the data sharing with other community members.
Moreover, the application presents also CCH Statistics, in terms of data quality, botnet impact and cyber events.

The community portal was deployed in April 2014, and is currently used by ACDC partners and community members. Therefore, the release of ACDC Analytics was postponed to early 2015, to give the consortium and the community a timeframe for the adoption of the portal in ACDC activities. The document was resubmitted on November 2015, and therefore the work presented refers to the status of ACDC Analytics in November 2015. This latest version complements this status by further analysis of the results presented in section 5.4.

The work on ACDC Analytics reported in this document is being carried out in the context of Task 6.2 of WP6.

2. Overview of the link between WP5 and WP6 deliverables (M24 deadline)

The community approach deployed in WP6 forms an integral part of the dissemination plan and is considered as one of the implementation tools of dissemination addressed by WP5.
The strategy for the community approach of WP6 is to enable stakeholders beyond the ACDC partners to become involved in ACDC with different levels of involvement, whilst the role of WP5 is to support this strategy through concrete dissemination activities. As these WPs are strongly interconnected, the following table provides an overview of which deliverable provides what information.
This section is repeated in all WP5 and WP6 deliverables.

Deliverables	***What is in the deliverable?***
D6.1.1 – user profiles and categorization	The different attributes used to categorize stakeholders, easing the prioritisation of the outreach activity of WP6 and the analysis of the different groups contributing to creating the ACDC community.
D6.1.2 – identified users list	The analysis of the stakeholders identified through different activities. This analysis is based on contacts established with 90% of the 426 identified stakeholders.
D6.2.1 – ACDC social platform	The description of the ACDC platform and the extension of its functionalities with respect to the original role foreseen in the DoW.
D6.2.2 – Adding social analytics to ACDC social platform	The addition of tools in the ACDC platform to monitor the activities and create a statistical overview of user activities
D6.3.1 – Involvement model for users in ACDC	A detailed description of the different activities that users can choose to be involved in ACDC, presented a UML graphs.
D6.3.2 – Report on user activities	A list of the activities carried out by ACDC partners over the first 12 months (updated on M20) of existence to lead to user involvement. Next steps identify the different activities proposed to users to become involve in ACDC; these activities are supported by the detailed approach in D6.3.1.
D6.3.3 – Report on users communities activities	The report presents the steps and actions proposed to users to become involve in ACDC and monitors their involvement from M12 to M24
D5.1.1 – Dissemination plan	The full list of activities defined to create awareness about ACDC and support the outreach activities of WP6
D5.1.2 – Intermediate dissemination report	The report of the dissemination activities of the first 12 months; this report is complemented by D6.3.2 for the section on individual meetings with organisations to reach the first level of involvement, i.e. letters of interest.

D5.1.3 – Intermediate dissemination report	The report of the dissemination activities of the second reporting year; this report is complemented by D6.3.2 and D6.3.3 for the section on individual meetings with organisations to reach the first level of involvement, i.e. letters of interest.

Table 1 – Overview of the WP5 – WP6 deliverables over the 30 months of operation

3. Introduction

The creation of a stakeholders' community is a major contribution to the out-reach approach of the ACDC project in fighting botnets across Europe. The creation of the ACDC community followed two connected lines of action. The first line focused on the identification of potential stakeholders, the definition of their relevance with respect to cyber-security in general, and to the botnet topic in particular, and the initial contact (identified as one of the ACDC partners) to invite them to join the community. The second line of action focused on the definition of the community governance, and on the creation of the community platform, used as the main channel for community members to interoperate. The community platform implements the ACDC governance, a key-structuring phase to ensure that interactions operate within a given set of shared rules.

Therefore, the ACDC community portal is a support element to facilitate the sharing of knowledge and the active participation of ACDC project partners and ACDC community members to activities proposed by ACDC. The ACDC user community portal has been moved from an "informative only" tool to the "one stop shop" service to access the ACDC infrastructure. This change included the implementation of the Data Access Manager (DAM) application, detailed in D6.2.1, as part of the community portal. Essentially, DAM enables community members to manage their API-Keys to the Central Clearing House (CCH), as well as to select the sharing policies for data shared within the ACDC user community.

The ACDC Community Portal was designed to:

- Provide a single front-end to access the ACDC Central data Clearing House (CCH)
- Foster the visibility (and uptake) of activities for the ACDC community
- Support interaction with stakeholders
- Enable fast and effective information sharing among members by providing different functionalities and different interaction areas.

The ACDC community portal is structured in two environments, one that is public and accessible to everybody and one that is restricted to ACDC members only.
The Portal is organized across different sections and provides a wide spectrum of online collaboration tools with the objective to share information and knowledge, improve communication between other stakeholders all over the Europe, share data and participate to experiments. These elements are described in D6.2.1, D6.3.1, D6.3.2, D6.3.3 and D6.3.4.

In order to support the ACDC user community in their interaction and participation mode, a set of social analytics tools were developed and integrated into the social platform with the aim of providing aggregated information and visual insights on managed user groups and profiles.
The purpose of the ACDC Analytics is to provide managers of the ACDC user community with insights on the community, to help them better understand the community needs and improve their decision making with respect to the community administration and animation.

4. Social Analytics Overview

Social analytics is a philosophical perspective developed since the early 1980s by the Danish idea historian and philosopher Lars-Henrik Schmidt. It differs from traditional philosophy as well as sociology, and it might be said that the perspective attempts to articulate the contentions between philosophy and sociology. The practice of social analytics is to report on tendencies of the times. (Schmidt, 1996 [1])

In particular nowadays, **social analytics measure the impact of social media on business**. It is an evolving business discipline used to understand how different kinds of conversations in online communities, including social networks, influence business performance (conversations driven by industry experts, competitors, prospects, customers, etc.). Increasingly, the ability to make sense of the social media conversational stream on Twitter, Facebook, YouTube and elsewhere on the Internet plays a crucial role in business.

Analytics experts have developed consensus around the definition of social analytics.

According to Gartner (2013) social analytics is defined as "the process of collecting, measuring, analysing and interpreting the results of interactions and associations among people topics and ideas. Interactions occur in workplace and external-facing communities. Social analytics is an umbrella term that includes a number of specialized analysis techniques, such as social filtering, social network analysis, influencer identification, profiling and scoring, social channel analysis, sentiment analysis and social media analytics, natural-language processing, and advanced techniques such as text analysis, predictive modelling and recommendations, and automated identification and classification of subject, topic, people or content. (Gartner, 2013 [18])

In particular, considering a business perspective, Jim Sterne, the author of Social Media Metrics, defines social media analytics as "the study of social media metrics that help drive business strategy". Jonas Klit Nielsen, managing partner at Mindjumpers, one of the premier social marketing agencies in Europe, agrees and adds that social analytics is "the ability to analyse performance of social media initiatives and social data for business intelligence". Besides Awareness Inc. defines social media analytics as "An evolving business discipline that aggregates and analyses online conversations (industry, competitive, prospect, consumer, customer) and social activity across social channels. Social analytics enable organizations to act on the derived intelligence for business results, improving brand awareness and reputation, marketing and sales effectiveness, and customer satisfaction and advocacy". (Awareness, 2012 [15])

Considering social analytics as the practice of gathering data from blogs and social media websites and analysing that data to make business decisions, different types of analysis can be considered. Qualitative analysis of social media is widely regarded to matter the most, but quantitative analysis is also gaining a great deal of attention now, too, especially with the rise of new social analytics tools.

Because of the importance and relevance of social analytics, the deliverable starts with a comprehensive overview of the domain, including

a) what is the main purpose of social analytics
b) What are the different types of input data used for analysis, which are the techniques used to analyse this data and the various methods for result representations
c) Which are the most popular social analytics tools

4.1. Social Analytics purposes

In social analytics it is important to determine which goals are addressed by gathering and analysing data. Typical objectives can vary according to the domain of applications and can include increasing

revenues, reducing customer service costs, getting feedback on products and services and improving public opinion of a particular product or business division. (Rouse, 2012 [22])
Specifically, purposes connected to social analytics implementation include:

- **Social goal**, relevant to the community animator point of view. This consists in measuring participation and contribution of users in a community, as well as popularities of contents with the purpose to enlarging the community and increasing the interactions among the community participants.
- **Reputational goal,** relevant to a personal point of view. This is fundamental in analysing the social interactions of a user with other people to calculate the reputation and the influence the user has online, in the context of one or more communities.
- **Economic goal,** referring to a business perspective. This goal deals with measuring the economic effects of user interactions with the system in order to increase the revenues derived from online interactions or communications.
- **Improvement goal,** also referring to a business perspective. This is related to the analysis of the huge volume of information contained in blogs, comments, forums, regarding products and services; this information includes consumer sentiments that can be used to evaluate users' experience with a particular product or service and can then be used to help companies perform better.

In order to achieve these goals it is important to define which data is needed, how it has to be examined and represented.

4.2. Kind of Analyses & related inputs

Social analytics has evolved from simple buzz and brand monitoring to the aggregation, segmentation and analysis of streams of social and web data to drive actionable business insight.
The first step in social analytics is to figure out how to monitor what happens in social media by collecting data generated in the social media channels. The question is “what data should be collected?”; for example it could be interesting to track every "mention" of a particular subject on social networks, or classify all those "mentions" as either "positive" or "negative" through some kind of "sentiment" analysis. In general these are different “metrics” that depend on the objectives of the analysis conducted (i.e. marketing products and services, influencer analysis, competitor monitoring, etc.).
Social metrics or quantitative measurement tools for social media are increasingly a component of the larger, related field of social analytics. In this context it is important to determine which are the most relevant inputs used for social analysis.
Some of the most common items or activities gathered to measure the impacts of social media are:

- **Number of visitors** to a website or web service
- **Referral sources** for website traffic, or where the traffic originated
- **Number of participants** in a network, such as followers, friends, fans, members
- **Amount of time** a participant spends engaging with a social network or website
- **Frequency of engagement**, or how often people interact with a brand, product or company (like, comment, share, post click, tag, etc.)
- **Positive or negative nature** of communication involving a brand, product or company; sometimes called sentiment analysis,
- **Volume of comments** of posts mentioning a particular service or company name,
- **Number of messages** sent about a particular brand or service,
- **Interaction rate**, or the percentage of total users of a service or social network who actively engage with it. (Walker, 2013 [25])

It is evident that social (media, technologies, applications) continues to penetrate the corporate world, so each use of social media generates and leverages different types of data, from simple

structured data about the characteristics of a brand's fan base, to unstructured (and often messy and noisy) postings in activity streams, to more complex data captured in a social graph. Some of this data (e.g. tweets, likes) is very *fast* with a short "shelf life," while other data is more enduring, such as a brand's social graph. All this creates *big data* and must be approached as such. Collecting, organizing and preparing this data for analysis can also be challenging.
In this context, the emergence of internet-based social media challenges not only data selection and collection, but also analytical techniques for data analysis. Compared to the data gathered through traditional means (e.g., experiments, surveys and interviews), social media listed data include much richer content (e.g., text messages, photos, videos, etc., generated by millions of users), and metadata ("data about data"). Given the richness and the high volume of social media data, appropriate analytical techniques are crucial for valid and accurate data assessments. Conducting data analyses with inappropriate analytical techniques lead to sub-optimal data assessments and / or incomplete information.
Currently, many analytical techniques (such as content and semantic analysis, social network analysis) have been adopted to evaluate specific aspects of social media data; these methods assist in detecting patterns in the content or the social relationships among social media users. The analytical techniques explore the important interactions among users, as well as the connections between content, the metadata, and users, and provide a dynamic perspective on data analysis, which includes the adaptation of dynamic network analyses, dynamic graph algorithms, and visual analytics. (Wang, 2012 [27]).
Because of their unique features, social media data pose a challenge to the analytic techniques used in traditional research; some of them are described below.

Trend Analysis: this is the practice of collecting information and attempting to spot a pattern, or *trend*, in the information. In social fields of study, the term "trend analysis" has many applications in order to understand changes in social patterns, and can be applied to various types of data. One of the most common examples of trend analysis is the identification of *emerging topics*. In order to really track the pulse of the blogosphere, it is important to detect phenomena that are only visible at the macro-level of the universe of relevant blogs. Given the plethora of posts being produced continually it is challenging but very valuable to determine large-scale patterns of what people are blogging about, and detect emerging areas of discussion. One commonly used approach to capture the notion of hot topics is to identify frequently occurring key-phrases; there has been considerable work in Natural Language Processing in identifying such commonly occurring collocation of consecutive words. (Melville, Sindhwani, and Lawrence, 2009 [19])
Trend analysis can also be applied to tags (*tag analysis*) to obtain a comprehensive vision of the most common tags cited by users, to the engagement rate in order to evaluate the participation about a particular subject, to fans/followers in order to establish the popularity of a particular product/person/subject, to the conversion rate (that aims to track the proportion of visits to a website who take action to go beyond a casual content view or website visit, as a result of subtle or direct requests from marketers, advertisers, and content creators).

Real time analysis: this is the process of delivering information about business operations as they occur; it means near to zero latency and access to information whenever it is required. This analysis provides a quick, at-a-glance overview of key metrics for bloggers and / or content marketers.

Sentiment and **Opinion mining analysis** refer to the use of natural language processing (NLP), text analysis and computational linguistics to identify and extract subjective information in source materials. The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market

their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and actioning it appropriately, the importance of sentiment analysis is increasingly relevant. (Barber, 2010 [16])
These analyses relate to the measurement of a consumer's attitude towards a brand, a particular product or service offered by the organisation and its competitors. It can be used to identify certain trends (e.g., a brand's loyalty), but more importantly, it can be used to make predictions, such as predicting the broad sales performance of a product (and if combined with location-based and channel data, it can predict the performance around a particular geographic location, or predict the sales performance of the social channel in comparison to the sales performance of other channels, such as email). (Melville, Sindhwani, and Lawrence, 2009 [19])
Some examples of these analyses include: brand awareness and reputation monitoring, polarity analysis, sentiment by company/competitor, etc.

Social graph analysis refers to the analysis of the graph's structure to reach certain conclusions. A graph's structure consists of the data associated with each *node* and the *connections/links* between each pair of nodes. In a social graph a node may represent an individual, or an organization; connections may represent certain relations about the nodes. This is a more recent form of social analytics and is made possible by the API-based accessibility of the various social graphs (e.g. Facebook's Open Graph). For example graph analytics are very often being used to identify a brand's *advocates* and establish which of these advocates are real *influencers*. *Influencer analysis* is a particular diffuse type of evaluation; having identified a subset of relevant blogs, it is useful to determine the most authoritative bloggers in this space. These are the experts or mavens whose opinions catch on most rapidly. It is important to identify this set of bloggers, since any negative sentiment they express could spread far and wide. These are the bloggers who marketers should keep most accurately informed and actively engaged. In addition to authorities, there are bloggers who are very well connected, who are mostly responsible for the spread of information in the blogosphere. These influencers are often (but not always) blogs with high authority. When presented with a large number of posts relevant to a topic, ordering them by the blogger's influence assists in information triage, given that it is not feasible to read all posts. (Melville, Sindhwani, and Lawrence, 2009)
Current systems are able to capture these activities on 'interaction spaces' (e.g. blog, wall, forum); by analysing the characteristics of these influencers (e.g., determine each influencer's lifetime value), marketers can treat them differently than other advocates and fans.

Geospatial analysis is an approach to apply statistical analysis and other informational techniques to data related to geographical or geospatial aspects.
It is used if the objective is analysing a company distributed around more cities' region, world's area or the globe. It shows a particular measure in one or more countries as compared to the whole continent (or the world) or helping in identifying, mapping, and analysing phenomenon. For example, it allows to see how a specific country is contributing to a website at a glance.
The social media feeds share a common nature: they are real time published expressions of a society's cultural and societal interests. Thus harvesting and analysing their content can offer unparalleled insight on socio-cultural dynamics. For example, they allow to:

- map the manner in which ideas and information propagate in a society, information that can be used for example to identify appropriate strategies for information dissemination;
- map people's opinions and reaction on specific topics and current events, thus improving the ability to collect precise cultural, political, economic and health data, and to do so at near real time rates; and identify emerging socio-cultural hotspots.

Geolocation information can be provided directly by the contributing users, if they decide to make this information available in user profile (static way), or it can be deduced from IP addresses if an IP geolocation solution is available (dynamic way).
By harvesting geographic content from social media feeds it is possible to transfer the extracted knowledge from the amorphous cyberspace to the geographic space, and gain a unique understanding of the human landscape, its structure and organization, and its evolution over time. (Stefanidis et al., 2012 [23])
This kind of analysis could be performed in order to observe measures over geographical aspect (conversion rate, visit by countries, revenue by countries, etc.).

The analysis of social data, when done correctly, can yield significant results; after having determined which social data needs to be collected and to be analysed, it is necessary to delineate how to visualize these results in order to improve their use for goal achievement.

4.3. Output Representations

The analytical techniques described allow analysts to integrate social media input data and analyze them; different kinds of representations can be used to summarize the results, so users can usually choose among various dashboards according to their needs.
"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance" (Few, 2007 [17]).
Graphical dashboards show a summary of social media activity and output of traffic generated; the most common output representations are:

Timeline graph: the timeline tells what is going on in the last "x" days. It permits to focus on specific days during a specified period that show a positive or negative trend in order to see what happened and what has been done during those days, in order to repeat it or avoid it (i.e. if in a timeline with two metrics -daily visit and revenue - there is a day during that period with a lot of additional revenue, it would be interesting to see what has been done that day and repeat it every week).

Figure 1 - Example of Timeline graph

GEO Map: The GEO map is a handy tool if a company, or more generally a community of people, is distributed around the globe; it allows to see how a specific country is contributing to a website at a glance. It shows for example what is the e-commerce conversion rate (that is the proportion of visits to a website which take action to go beyond a casual content view or website visit) in a specific country as compared to the whole continent (or the world), or how this rate changed in two different time ranges. The GEO map can also show the real time activity either at country or at city level.

Figure 2 - Example of GEO Map

Table: Tables are used for monitoring landing pages, content, product and marketing campaign performance. It allows to see a list of the dimension chosen along with metrics chosen. For example, considering a blog on a website, a table can be very useful to understand if certain posts convert better than others showing landing page in a column and goal completions in another.

Ecommerce Conversion Rate

Medium	Visits	Ecommerce Conversion Rate
organic	580,543	5.76%
affiliate	501,281	3.15%
(none)	331,764	9.31%

Figure 3 - Example of Table

Pie chart: Pie charts can be used to visualize data such as percentages out of a whole. They are widely used in businesses contexts. For example, through such as this chart it is possible to see how much of branded traffic comes to paid and organic search campaigns.

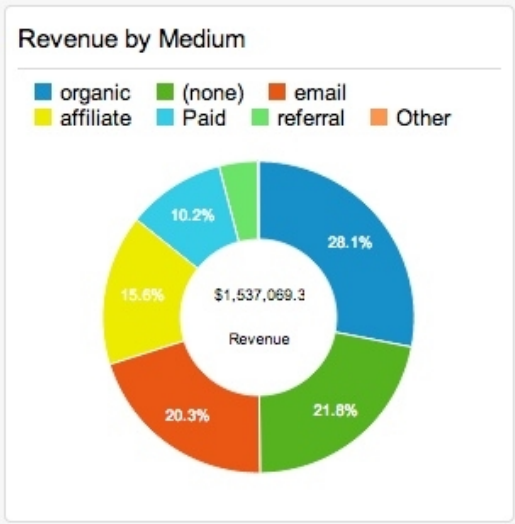


Figure 4 - Example of Pie chart

Bar chart: Bar chart is probably the most powerful of all visualisations because it offers a multitude of customizations options. There are a variety of bar charts that can be created, including vertical, horizontal, grouped, stacked, and overlapped versions of each. It is possible to create charts where showing a metric - on the y axis (i.e. Revenue), group it by a certain dimension - on the x axis (i.e. Medium) and pivot it by a second dimension - different colour of a bar (i.e. a subcontinent). The potential of this chart is that it tells a story; bar charts are good for side-by-side comparison and spotting trends in a small number of discrete data points.

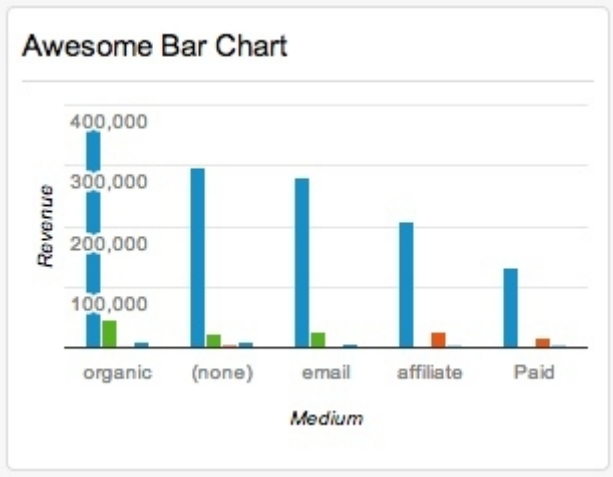


Figure 5 - Example of Bar chart

Heat Map: Heat Map is a graphical overlay of a website which points out what content is hot and what not. This is mainly done by tracking the mouse clicks or using eye tracking of the visitor. The click-heat map makes it possible to analyse the clicking behaviour of the visitor while the eye tracking-heat map shows where people actual look on a webpage. A heat map gives the opportunity to look through the eyes of visitors as they navigate on the website (i.e. the heat map shows that an advertisement is easily overlooked because of the wrong colour or because it is close to big images which capture more attention). If the site has a dynamic content, it is more difficult to take the full advantage of heat maps because people click on different places if content changes. This type of representation is not very common and it can be provided by some tools such as Crazyeggs, Clickheat, WordPress plugin #1, Mouseflow, Clickdensity, etc.

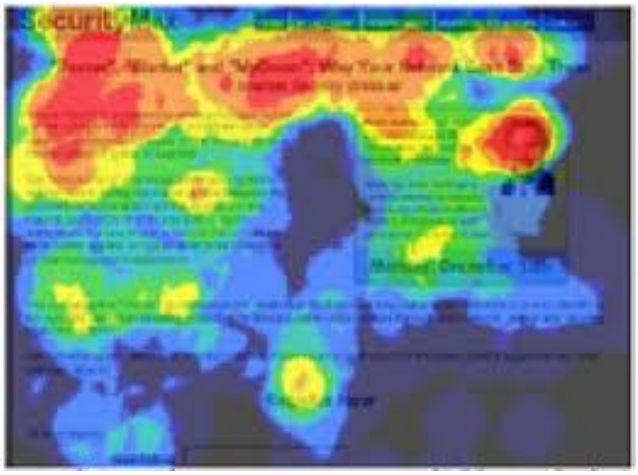


Figure 6 - Example of Heat map

Once the output representation is built, the different tools make available some functions of sharing, exporting and customizing the dashboards:

- sharing is a functionality which allows to share dashboard created with teammates;
- export is a functionality which enables to download a PDF version of the dashboard;
- customize dashboard is a functionality which allows to decide the layout of the dashboard (i.e. making graphs larger than others). (Waisberg, 2013 [24]; Olyslager, 2012 [21]; Noble, 2012 [20])

Some of the tools used for analyzing unstructured data found in different Internet resources and for showing structured results will be described in the next paragraph.

4.4. Rise of social analytics tools

There is a plethora of social media monitoring and measuring tools on the market today to accomplish all of the goals described.

Popular software programs that allow people and companies to listen to and analyze content in social networks include Sysomos, and Salesforce-owned Radian6. There are hundreds of other tools, including many that are being combined into unified software suites to simplify the work.

Since 2011 in particular, big software and Internet companies have been showing more interest in buying or partnering with social monitoring and measurement start-ups. The wave started in 2011 when online business software company Salesforce bought one of the leading firms in this field, Radian6. In mid-2012, Oracle bought three social metrics-oriented firms (Involver, Collective Intent and Vitrue) and disclosed plans to integrate them into a "unified social platform" that would allow businesses to analyse behaviour and content across many social networks from a single interface. Google bought Wildfire, a social marketing and advertising service, in July 2012. A month later, Facebook announced it was buying Threadsy, the creator of a social media influence measurement tool called Swaylo.

Choosing Tools by Function

A plethora of other tools exist to help with social analytics. The choices of which to use are typically governed by the business (or other) goals someone wants to achieve, along with the size and budget of the organization.

Enterprise software giants such as Oracle and SAP offer more expensive, but also powerful tools for monitoring, measuring, analysing and engaging on social media. Increasingly, these activities are

being unified into integrated software suites to make data handoff easier and to simplify social media management for those doing it.
These large, integrated enterprise software suites tend to be too expensive for many individuals, sole proprietors and small businesses. But other less expensive and free options are available.
Tools are often classified by their function, including:

- **Social media monitoring:** These tools scour conversations occurring in social media and typically look for the keywords and phrases of interest; for that reason, they're called "listening" tools. Businesses use them to track how often their products, services and company names are mentioned, along with those of their competitors. Another common use is simply to track trends in a particular field or industry. Focus is on identifying keywords that people are using in conversations that seem relevant to particular business goals.
- **Social media measurement:** These tools also do monitoring but add the capability of measuring and refining the data-crunching based on filters and parameters. Many of these social metrics apps are specific to a type of social media, such as Twitter or Facebook. Knowing what to measure can be tricky and involve anything from simple mentions to more complex social media influence tracking.
- **Social media engagement:** These tools allow or help automate communication across various social media platforms and they often are called "engagement tools," "social media management" tools, or simply "social media dashboards". The key function of "engagement" tools is communication, meaning that they allow someone to log into a single software program and then post updates or comments to various social platforms, along with photos and videos.

The most powerful social analytics platforms combine all three steps of engagement, monitoring and measurement into a single application; some enterprise examples are Salesforce, Oracle and SAP, described in the section below. (Walker, 2013 [25])

The most popular tools
In 2013 Gartner conducted a survey in order to profile social analytics solutions that cover social filtering, sentiment analysis and public-facing social media analysis. The report focuses on solutions that are at least 60% software-based, including SaaS, as opposed to solutions that are mainly based on consulting services.
With some exceptions, the social analytics vendors profiled can analyse the usual information sources, such as, Twitter (@mentions, #tags, keywords and Twitter-specific handles), Facebook pages and public data, YouTube, blogs, forums, and news stories, with the following exceptions:

- IBM and Opera Solutions do not analyse Facebook pages;
- Socialbakers does not analyse Twitter #tags, blogs, forums or news stories.

For other, less commonly examined information sources, vendors' analysis capabilities vary, often because of the limitations of social networks' APIs or concerns about violations of terms of service. In particular Table 2 details which tools supports the less commonly examined information sources. (Gartner, 2013)

Vendor	LinkedIn Company Pages	LinkedIn Groups	Google+	Pinterest	Klout Score	Flickr	Social Data Aggregators
Adobel	x	x	x			x	x
GoodData	x	x	x	x	x		x
IBM Social Media Analytics		x	x		x		x
Oracle		x	x		x		x
Salesforce.com	x	x	x	x	x		x

(Radian6)							
Socialbakers			x				
Sysomos (Marketwired)	x	x	x	x		x	x

Table 2 - Capabilities of Profiled Social Analytics Vendors in Relation to Less Commonly Examined Information Sources

These are the most popular tools proposed by vendors in Table 2.

Adobe Social

Adobe Social is the solution proposed by Adobe, that entered the social analytics space in 2011 to broaden its portfolio of marketing applications, and in 2012 acquired additional social media application technology from Efficient Frontier, which had earlier acquired Context Optional. The product measures and reports on over 100 different engagement metrics across Facebook, Twitter, Google+ and LinkedIn. It tracks post-performance to provide insight into how a specific piece of content, time of day or week drives conversations, and it identifies key influencers. Adobe Social is an interesting product because of its integration with Adobe Analytics (Omniture). Together, they can measure the impact of social posts on website activity and behaviour including click-throughs, page views, time on site, leads, revenue and other relevant conversions. Moreover, predictive algorithms within the Adobe Social publishing tool can predict post performance and recommend changes to post attributes to improve engagement.

Key facts:

- Geographies: Most Adobe Social customers are large enterprises in North America that need to manage multiple social profiles spanning regions, departments and individual brands across the company.
- Languages: The product supports 17 languages for listening, publishing and moderation; the user interface (dashboards) is rendered in nine languages, and sentiment analysis is supported in nine languages.
- Reporting: Adobe provides out-of-the-box reports for measuring brand, engagement, and conversion metrics, as well as administration activity (total posts published, moderated, post escalations and so on). The majority of reports can be customized (dates, filters and so on) and all can be exported to comma separated values (CSV) files with limited alerting capabilities.
- Integration: Adobe's technology partner program draws on API or custom development. Currently, most CRM integrations are custom.
- Industries: Adobe Social is deployed across a range of industries, but its customers are concentrated in media and entertainment, retail, automotive, transportation, and travel and hospitality.

GoodData solution

GoodData is a privately held company based in San Francisco which offers an end-to-end, cloud-based Business Intelligence (BI) and data warehouse platform with integrated social analytics. What is interesting about GoodData is that it is not just a social analytics platform; it is an analytics platform that can use built-in NLP capabilities, along with sentiment and influencer analysis capabilities to combine, analyse and correlate a broad range of multi-structured data sources, derive insights across social and enterprise data, and enhance any analytic process. GoodData is best suited for customers needing customized and complex analytical dashboards that combine multiple data sources (for example, combining social data with proprietary point-of-sale data).

Key facts:

- Geographies: The majority of GoodData's customers are in the U.S. It also has a small number of customers in Europe and other countries.
- Languages: The platform supports UTF-8 unicode characters throughout, so all user-generated content, such as data loaded (data ingestion) and names of reports, dashboards and metrics, can be in any language. GoodData is working on localizing its user interface (navigation, menus and so on) into 13 languages initially, and plans to release this support early in 2014.
- Reporting: GoodData offers an extensive set of reporting, dashboards, and interactive analysis and visualization platform capabilities, as well as scheduling distribution and alerting that can be used to create social analytics reports.
- Integrations: GoodData's data source integration includes external social networks (such as Facebook, Twitter, Pinterest and YouTube), internal social applications (Yammer and Box),social media engagement solutions (such as HootSuite and Scup), customer databases (such as salesforce.com and Microsoft Dynamics), data aggregators (such as Gnip and DataSift), and other sources (such as Google Analytics and Adobe Analytics), along with APIs to campaign management and other enterprise software vendors.
- Industries: GoodData's BI customers span many industries; its social analytics customers are typically advertising agencies or retail organizations.

IBM Social Media Analytics

The IBM Social Media Analytics solution is available on premises or as a SaaS-based offering. It is an evolution of the IBM Cognos Consumer Insight product. The analysis refreshes every 20 minutes, making the solution sufficiently real-time for market research but not adequate for crisis or customer service management. IBM Social Media Analytics' integration with IBM Cognos Business Intelligence 10.1 makes it an attractive investment for organizations looking to incorporate social analytics into a broader enterprise BI strategy. The solution can algorithmically identify topics based on sets of words that are frequently used together and compare topic activity side by side. IBM Social Media Analytics can also perform affinity analysis to identify and communicate the strengths of relationships within a social network.

Key facts:

- Geographies: The majority of IBM Social Media Analytics customers are based in North America, but IBM's customer base is expanding across the globe to Europe and Asia.
- Languages: Social text can be analyzed in any language. Sentiment analysis is available for key languages, including English, French, German, Dutch, Spanish and Chinese (traditional and simplified).
- Reporting: The IBM Social Media Analytics reporting dashboard is divided into three sections: share of voice, segmentation and discovery. All three allow for deeper views into verbatim text, the ability to interact with the data, and the ability to export data into a third-party application.
- Integration: IBM Social Media Analytics offers seamless integration with a client's external data and broader analytics dashboard. For predictive and deeper text analytics, this is achieved most easily with IBM SPSS, and for custom dashboards and reports with IBM Cognos Business Intelligence 10.1, which serves as the application's front end.
- Industries: IBM Social Media Analytics is strong in IBM's base markets: telecommunications, banking, publishing, government, consumer products, retail, pharmaceuticals, media and entertainment.

Oracle Social Relationship Management

Oracle entered the social relationship management market in 2012 with the acquisition of three companies: Vitrue, a social marketing and publishing platform, Collective Intellect, a social

intelligence firm, and Involver, a company that enables point-to-point engagement. Oracle Social Relationship Management (SRM), the result of these acquisitions, enables social listening, social engagement, social publishing, content and app creation, and social analytics from one interface. Customers use Oracle SRM to conduct social analysis that is unified across earned, owned and paid-for social media. This includes measuring fans' growth, engagement, reach, as well as consumers' intentions with regard to purchasing, complaining, switching, loyalty and so on. What is interesting about Oracle SRM is its combined use of keyword search and latent semantic analysis to expose latent contextual meaning within a large body of text. It does this by looking at word usage (specifically, co-occurrence of words) within a set of documents. Latent semantic analysis groups similar conversations into clusters. Users can save a search as a topic to reuse later.

Key facts:

- Geographies: Most of Oracle's customers are in the U.S., but there is growth in EMEA, Latin America and Asia/Pacific.
- Languages: The Oracle SRM user interface is available in 30 languages, and there are plans to support additional ones. Monitoring is supported in 9 languages (English, simplified Chinese, Portuguese, Spanish, German, Italian, French, Japanese and Korean) and there are plans to support 17 by early 2014.
- Reporting: Oracle SRM offers real-time reports and interactive dashboards plus drill-to-detail capability and email alerting. Visualizations for metrics such as daily activity, voice share, sentiment and source distribution are also provided. Reports are available for competitive analysis, content optimization, campaign and post-campaign analysis, brand analysis and crisis reporting. Reports and dashboards come with 19 out-of-the-box filters called indicators that categorize conversations by contextual meaning to identify signals for further analysis or engagement. Indicators can be customized to align with a company's key performance indicators and marketing strategies or to reflect language used in specific industries or by specific audiences. Users can request reports with custom metrics through a services engagement.
- Integration: Oracle SRM provides initial levels of integration with various Oracle applications, including Oracle RightNow Cloud Service, Oracle CRM On Demand, Oracle Sales Cloud, Oracle Siebel CRM, Oracle Eloqua Marketing Cloud Service and Oracle ATG. Integration with Oracle Endeca and Oracle Business Intelligence Enterprise Edition is planned. Integrations with other enterprise applications are through the use of APIs and Web services.
- Industries: Oracle customers span all industries, but there is a concentration in the consumer goods, media, restaurant, technology, financial services, travel and transportation, beverage and retail sectors.

Radian6 (salesforce.com)

Radian6, acquired by salesforce.com in 2011, serves as a core component of salesforce.com's Marketing Cloud solution, which provides Marketing Cloud with social media monitoring and analytics capabilities. Radian6 is offered in a SaaS-model and is best suited to companies upgrading from "freemium" social solutions aiming to become acquainted with social media analysis. It is also of interest to midsize and large organizations with multiple stakeholders that need to manage engagement on multiple social channels. The Radian6 Analysis Dashboard consists of a series of independent widgets that can be drilled down to verbatim text or the "river of news." Radian6 itself is a base social media analytics solution with a partner program, Insights, which enables users to take a deeper look into complex social analyses of factors such as emotion, intent to buy and communities of influence.

Key facts:

- Geographies: Most Radian6 customers are in North America, but there also clients in Europe and Asia.

- Languages: The solution can analyze over 23 languages.
- Reporting: The simplest way to consumer Radian6's data output is through the Radian6 Summary Dashboard. This includes a clean interface to view volume, sentiment, demographics and channel breakdowns, along with changes over time.
- Integration: Radian6 integrates with Google Analytics, Webtrends and Adobe SiteCatalyst to correlate Web traffic data back to social media posts.
- Industries: Radian6 is strongest in the professional services, high-tech, media, telecommunications, retail, consumer goods and transportation sectors.

Socialbakers (Analytics)

Socialbakers is a four-year old, SaaS-based, private company headquartered in Prague. It is primarily a social media benchmarking firm known for its analytics, one that also has an application for social media engagement and a professional services arm that provides strategic advice and analytics interpretation. Socialbakers has two key products focused on social analytics, Analytics Pro and Ad Analytics:

- Analytics Pro is known for competitive benchmarking and can measure interactions such as likes, comments and shares; response times; response rates; and "engagement rate," which is an average interaction measurement as opposed to a measurement per individual fan. This product also includes market benchmarking tool that can filter by region and identify top content within a demographic group.
- Ad Analytics is an ad analytics, optimization tool.

Socialbakers' solutions are best suited to companies looking to do competitive and market benchmarking alongside other social analytics tools. Their graphical representation of data is highly attractive and easy for business professionals without a market research or analytics background to understand.

Key facts:

- Geographies: Europe makes up the majority of Socialbakers' customer base, with Latin American, North American and South American customers coming at a close second. Socialbakers does about 15% of its business in Asia.
- Languages: The Socialbakers dashboard is in English, French and Russian, but can monitor all major global languages, including Korean and Arabic.
- Reporting: Socialbakers offers standard executive reports. These reports give a high-level overview of a company's social media activity and compare it with other companies in its industry. All data can be exported to PDF or CSV files.
- Integrations: Google Analytics, Net Promoter Score indexes and Facebook advertising campaigns. Customized solutions can be created for individual client needs.
- Industries: Socialbakers deals primarily with business-to-consumer organizations in the retail, airline, financial services and creative services industries.

Sysomos (Marketwired)

Sysomos offers a social media analytics platform. In 2010 it was acquired by Marketwire, which changed its name to Marketwired in 2013. It uses Sysomos to power its Media Analysis Platform (MAP) and Heartbeat products.

MAP researches, analyses and reports on real-time and historical social media conversations. It automatically determines sentiment and identifies key influencers by demographic or industry. Heartbeat monitors and measures social media conversations in real time and enables engagement. The products can be used separately or together. Sysomos also offers custom applications. Sysomos's offerings are interesting because the Sysomos analytics engine enables ad hoc searching within a two-year historical time frame. MAP and Heartbeat will be of interest to chief marketing officers and digital marketers who want to link to CRM applications using Sysomos's APIs.

Key facts:

- Geographies: Most of Sysomos's customers are based in North America. It also has a strong presence in Western Europe and does business in other parts of the world.
- Languages: 186 and translation from 64 languages.
- Reporting: Standard reports include dashboards and executive overviews, voice share and social integration analysis. Customers can create custom reports and export them to other applications.
- Integration: Available for any system that can plug into the REST-based API and accept the XML output.
- Industries: Agencies and industries that converse using social media.

5. Modelling ACDC Social Analytics

5.1. Purposes of the ACDC Analytics

The purpose of the ACDC Social Analytics is to provide managers of the ACDC user community with insights on the community, to help them better understand the community needs and improve their decision making with respect to community administration and animation.

As introduced in the previous WP6 deliverables, the ACDC user community portal evolved from an "informative only" tool to the "one stop shop" service to access the ACDC infrastructure [11]. This change included the implementation of the Data Access Manager (DAM) application, detailed in D6.2.1 [12], as part of the community portal. Essentially, DAM enables community members to manage their API-Keys to the Central Clearing House (CCH), as well as to select sharing policies for data shared within the ACDC user community.

Given the above evolution of the portal, the ACDC Social Analytics has been also structured to report the community activities at two levels: *user content* and *data sharing.*

As far as the user content level is concerned, the purpose of the ACDC Analytics is, in particular, to:

- **Improve portal design and services:** With the help of ACDC Analytics it is easy to find out how users interact with the portal and then improve design and functionalities;
- **Assess** the community growth in terms of members and contents;
- **Track mass user activity:** Using ACDC Analytics allows the managers to measure how users interact with specific resources on each page en masse. These analytics show how the resources are used and how interesting they are for the community, and how users, in general, respond to specific calls to action and messages;
- **Capture user data** to understand attitudes, opinions and trends inside ACDC Community Portal;
- **Reveal** how users connect and interact with each other;
- **Predict user needs** and **improve user satisfaction** by anticipating their needs and interests and therefore recommending next best actions;
- **Create customized campaigns** and **promotions** that resonate with the users of the ACDC Community Portal and respond to their interests;
- **Identify the primary influencers** within the ACDC Community Portal;
- **Understand the resources and topics** that are of great interest and which ones generate more interaction inside the ACDC Community Portal;

As far as the data sharing level is concerned, the purpose of the ACDC Social Analytics is to present statistics on the adoption and usage of the DAM application. This can be of help for the ACDC user community in many ways; among others:

- **assess the participation** of the ACDC user community to the data sharing activity, that is essential to the ACDC approach to fight botnets. This is particularly important, for instance, to correct lack of involvement in the data sharing from stakeholders of a given kind that are already members of the ACDC user community (CERT, ISP, Antivirus, etc.)
- by analysing the current distribution of data sharing stakeholders the community managers can also get information to **better planning of involvement campaigns**, with the aim of

involving a new kind of stakeholders to the data sharing tasks, or to increase the number of participants of an already existing kind.

- **envisage potential issues** related to the data sharing functionalities, by analysing the current usage of key and sharing policy management by different DAM users, and project them accordingly to the estimated growth of the ACDC user community.
- **understand sharing habits** for each kind of stakeholders. This may be of help for new stakeholders in the community, community managers can suggest them potential partners for data sharing, thus fostering their interaction with the existing community.

One important point about the ACDC analytics is that the *data sharing* level does not deal with the quantity and quality of the data exchanged through the CCH. This point, also covered by the community portal, has been implemented in the Botnet Metric section of the portal, developed in the context of WP4, and described as part of D4.4 – Publicly Accessible Database of Botnet Metrics [14].

Monitoring the activities of ACDC user community is the process of observing and recording information about users that relates to their activities on the Portal. Monitoring is not about watching every move that a user makes and recording their every action, but rather making sure that the community portal provides the necessary services at the right time and aligned to the needs of the users. It also helps anticipate changes in the interests of users and is a way of telling if a user is progressing or not in a positive way, and therefore take the correct decisions and actions.
Therefore, in terms of the user content, ACDC Analytics is a tool that measures, collects and analyses the data to understand the users on ACDC Community Portal ***only***, so as to **improve the user experience and to put forward the ACDC services.**

The above distinction between analytics on *user content* and *data sharing* has a strong impact on how the ACDC Analytics portlet has been conceived and developed, as described in the section below.

5.2. ACDC Analytic dashboard design

5.2.1. User contents

The ACDC Analytics developed for ACDC Community Portal, at a user content level, follows two different approaches: the **Stakeholders analysis approach** and the **Resources analysis approach**.

In the following, the term “Resource” indicates a single item of content stored by the ACDC community portal. Sample of Resources in the portal are news, messages in the forum, wiki pages, initiatives, stakeholder profiles, tools, etc. Each resource belongs to a specific type that defines the resource structure and visualization (e.g. initiatives are visualized in the Initiatives & Stakeholders application). Most of the resources in the portal can be subjects of social actions (samples of social actions on a resource are like, follow, comment, notify to other users, etc.). The set of social actions that can be performed on each resource depends on the type of the resource itself.

On the other hand, the term “Stakeholder” indicates a single organisation that joined the ACDC Community Portal. Inside the portal, each stakeholder has one or more users who interact on behalf of the stakeholder. Therefore, each social activity (e.g. add, comment, like etc. a resource) is done by a user on behalf of the stakeholder that he belongs to.

5.2.1.1. Stakeholder analysis approach

ACDC Community Portal is a platform oriented to a community grouping stakeholders that shares the same goal of fighting botnets at a European level. Internally, each stakeholder has a number of users who interact on the portal on behalf of its organisation. Therefore, the stakeholders' analysis approach focuses on the relationship between organisations and users. The added values of retrieving information at a stakeholder level are:

- analyse how animated the portal is based on the stakeholders activities;
- understand the level of interaction of each stakeholder - for instance by differentiating between the provisioning of new information (and knowledge), and the social interaction around the contents created by others;
- have a concrete map of who's doing what;
- observe the active and passive stakeholders in a concrete moment;
- compare the level of interaction of a stakeholder based on different periods of the year;
- have a top 10 stakeholders in terms of interaction;
- identify a negative or positive progress of a stakeholder during a defined period of time;

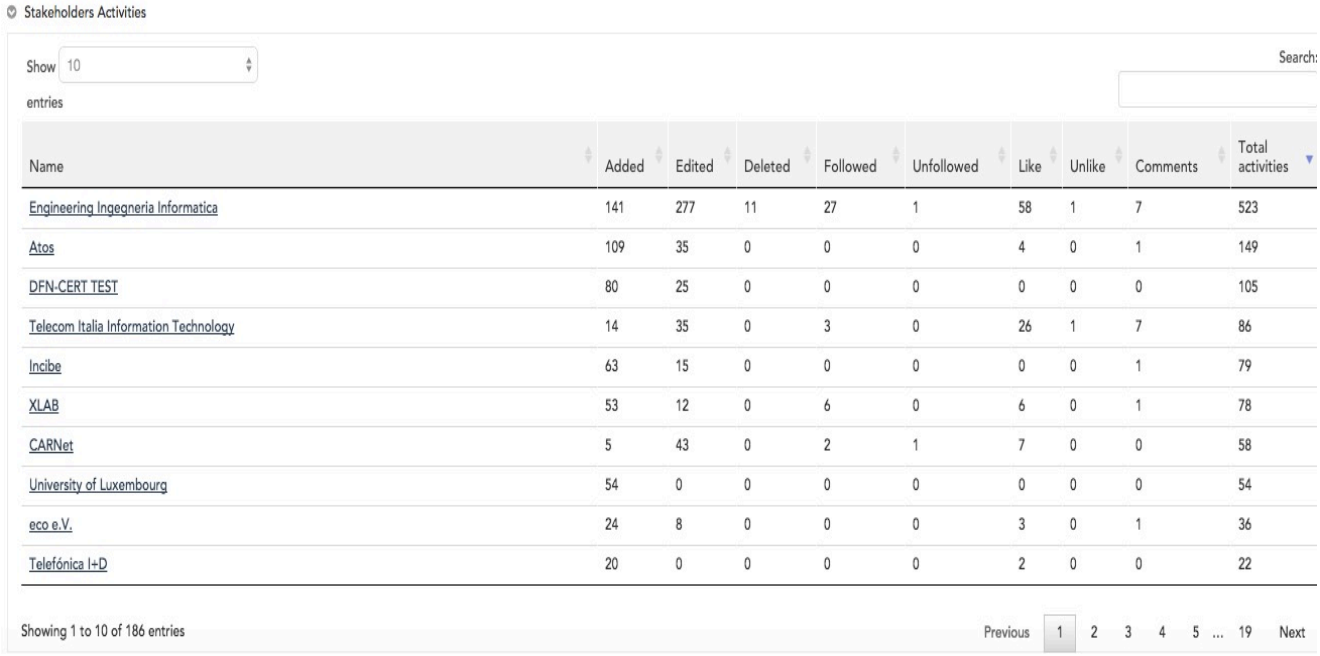
In concrete, ACDC Social analytics, at a stakeholder level, offer the following types of information (as it can be seen in the figure below):

- name of the stakeholder;
- total activities done by the stakeholder;
- number of login and logout operations;
- number of resources added, edited and deleted by the stakeholder;
- number of resources followed (and unfollowed) by the stakeholder;
- number of resources liked (and unliked) by the stakeholder;
- number of comments added to resources by the stakeholder;
- number of uploaded (and edited) documents by the stakeholder;

Based on the above types of information offered by the ACDC Social Analytics, managers can understand the stakeholders: i) who are passive actors (the stakeholders who login, but don't interact or have a low level of interaction); ii) who are active actors (the stakeholders who provide content and therefore knowledge by adding resources, documents, commenting the resources, etc.); iii) who are socially engaged without providing knowledge (stakeholders who don't provide content, but prefer to use the social functionalities "like", "follow", etc.).

Given the (potentially large) amount of information stored in the portal, the stakeholder analysis has been enriched with filtering and ordering functionalities, which gives the possibility to retrieve valuable information. In particular it is possible to:

- filter the data by date, which gives the possibility to manipulate the data on an interval of time, thus having a closer look, for instance, to changes in the activity level of a stakeholder over time;
- limit the analysis to a single stakeholder, by searching it in the search bar;
- sort from the biggest to smallest number for each kind of interaction listed above, or oldest to newest based on the needed data, thus revealing significant differences in the choices of the stakeholders with respect to specific kinds of interaction.



Stakeholders Activities

Show 10 entries

Search:

Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total activities
Engineering Ingegneria Informatica	141	277	11	27	1	58	1	7	523
Atos	109	35	0	0	0	4	0	1	149
DFN-CERT TEST	80	25	0	0	0	0	0	0	105
Telecom Italia Information Technology	14	35	0	3	0	26	1	7	86
Incibe	63	15	0	0	0	0	0	1	79
XLAB	53	12	0	6	0	6	0	1	78
CARNet	5	43	0	2	1	7	0	0	58
University of Luxembourg	54	0	0	0	0	0	0	0	54
eco e.V.	24	8	0	0	0	3	0	1	36
Telefónica I+D	20	0	0	0	0	2	0	0	22

Showing 1 to 10 of 186 entries

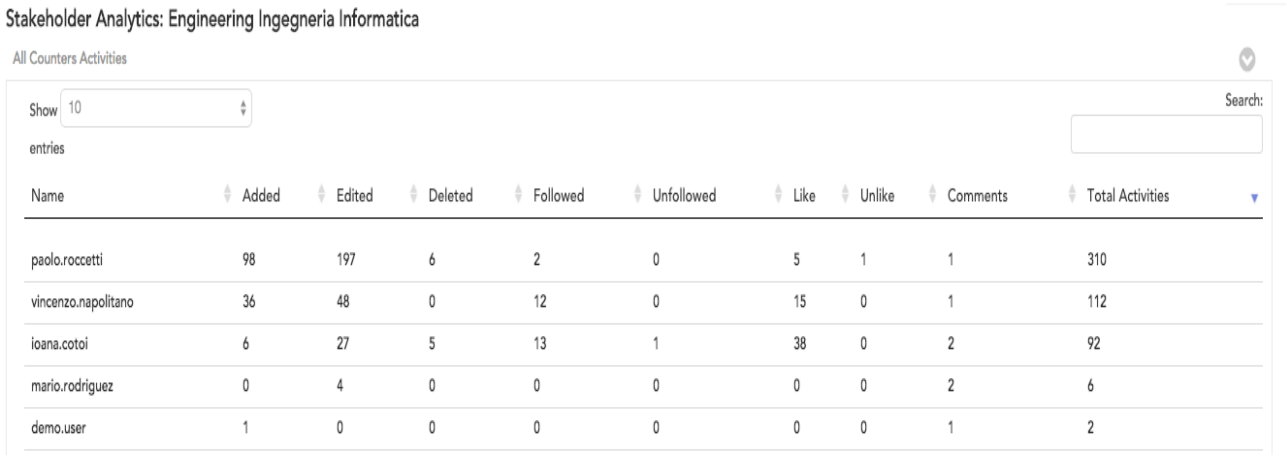
Previous 1 2 3 4 5 ... 19 Next

Figure 7 – Stakeholder activities, overall view

Moreover, ACDC Social analytics provides also detailed information on single stakeholder, meaning that by clicking on a stakeholder name, the manager will have a complete view of the types of activities done by the users belonging to that stakeholder, to better understand which users are making real, valuable contributions.

At this level the real activities done by the users are not visible, but it gives a good image of what happens inside an organisation. The ACDC Social Analytics provides the list of all the users that belong to a stakeholder and the total activities done by the stakeholder divided on single users.

These types of information can help the manager understand how the users interact over the time, understand what types of users a stakeholder have and take the correct decisions in order to better engage them into the services offered by ACDC Community Portal.



Stakeholder Analytics: Engineering Ingegneria Informatica

All Counters Activities

Show 10 entries

Search:

Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total Activities
paolo.roccetti	98	197	6	2	0	5	1	1	310
vincenzo.napolitano	36	48	0	12	0	15	0	1	112
ioana.cotoi	6	27	5	13	1	38	0	2	92
mario.rodriguez	0	4	0	0	0	0	0	2	6
demo.user	1	0	0	0	0	0	0	1	2

Figure 8 – Stakeholder activities, detailed view

5.2.1.2. *Resources analysis approach*

As mentioned before, the ACDC Community Portal aims to foster a wider level of information and data sharing and, therefore, a faster and more effective communication between stakeholders active in the cyber security area with the final aim to fighting botnets across Europe. In order to support active participation to ACDC, the ACDC community portal is organised not only to present information, but to allow social interaction around this information – allowing users to rate tools and solutions, allowing stakeholders to navigate "who's who" in cyber in Europe – and most of all allowing users to continuously enrich the knowledge that is presented.

Therefore, a set of features were developed and implemented on ACDC Community Portal, such as News, Initiatives and Stakeholders, Tools and Services, Documents, Experiments, Forum and DAM.
ACDC Social Analytics focuses also on analysing the resources related to each feature. The added values of retrieving information at resource level are:

- observe the evolution of the ACDC Community Portal in terms of activities;
- analyse how animated the portal is on a defined period of time;
- understand the usefulness of a feature;
- measure the effectiveness of a campaign;
- observe the lifecycle of a resource;
- find the top 10 resources per each feature;
- identify the negative or positive trends related to the features and resources;
- observe the resources which produce rumours;
- find the resources which pulled in the most social activities;
- understand how popular a resource is based on the number of visualisations;
- evaluate, in an effective manner, the impact of an action or a campaign;

In concrete, ACDC Social analytics, at the resource level, for each feature offer the following types of information (as it can be seen in the figure below):

- name/title of the resource;
- total activities done related to the resource;
- created date of the resource;
- number of visualisations of the resource;
- number of changes done to the resource (added, updated and deleted);
- number of followers (and unfollowers) of the resource;
- number of like (and unliked) of the resource;
- number of comments added to the resource;
- number of markers added or cancelled to the resource;
- numbers of categories added to the resource;
- numbers of tags added to the resource;

Total Activities	View	Added	Updated	Followed	Unfollowed	Like	Unlike	Commented
792	364	154	144	22	3	98	1	6

Show 10 entries

Search:

Name	Create Date	View	Updated	Followed	Unfollowed	Like	Unlike	Commented	Total Activities
botnet of infected PoS	2014-06-27 12:03:49.0	27	1	1	0	1	1	0	32
New way of Dridex banking malware infection	2015-03-30 12:18:35.0	11	1	1	0	1	0	0	15
Spam Campaign hits German Telekom - do you see something similar in your country?	2015-08-20 10:27:59.0	11	1	0	0	2	0	0	15
Phishing Education Can Save Nearly $4m Annually	2015-08-27 15:41:22.0	7	1	1	1	2	0	0	13
Google, Yahoo, Facebook Collaborate to Blacklist Bad Bots	2015-09-10 18:02:49.0	5	1	1	0	3	0	2	13
What is the Data Protection Authority?	2014-10-16 12:57:42.0	9	1	0	0	0	0	0	11
Operation Lotus Blossom: A New Nation-State Cyberthreat?	2015-07-07 15:31:02.0	7	1	0	0	2	0	0	11
"EUROPEAN INITIATIVE FOR FIGHTING BOTNETS – ADVANCED CYBER DEFENSE CENTRE" - after conference	2015-06-04 10:59:05.0	6	1	1	0	2	0	0	11
MixMod released	2015-04-08 14:28:20.0	5	1	1	0	3	0	0	11

Figure 9 – Resource activities, overall view

The types of information mentioned above offered by ACDC Social Analytics can be divided into 2 categories: information related to actions done by the creator of the resource and the information related to the actions done by the other users.
Given the (potentially large) amount of information stored in the portal, the resource analysis has been enriched with filtering and ordering functionalities (similar with those added for the stakeholders analysis), which gives the possibility to retrieve valuable information. In particular it is possible to:

- filter the information by date, which gives the possibility to manipulate the data on an time interval, thus taking a closer look, for instance, to changes in the activity level of a resource over time;
- limit the analysis to a single resource, by searching it in the search bar;
- sort from the biggest to smallest number for each kind of interaction listed above, or oldest to newest based on the needed data, thus comparing the resources based on different categories.

Moreover, by applying data filters, the ACDC Social Analytics provides the data requested on a dynamic chart. By adjusting date ranges, filters, and graphs on a dynamical way, managers can customize the performance data to see only what interest them, making campaign maintenance faster and easier.

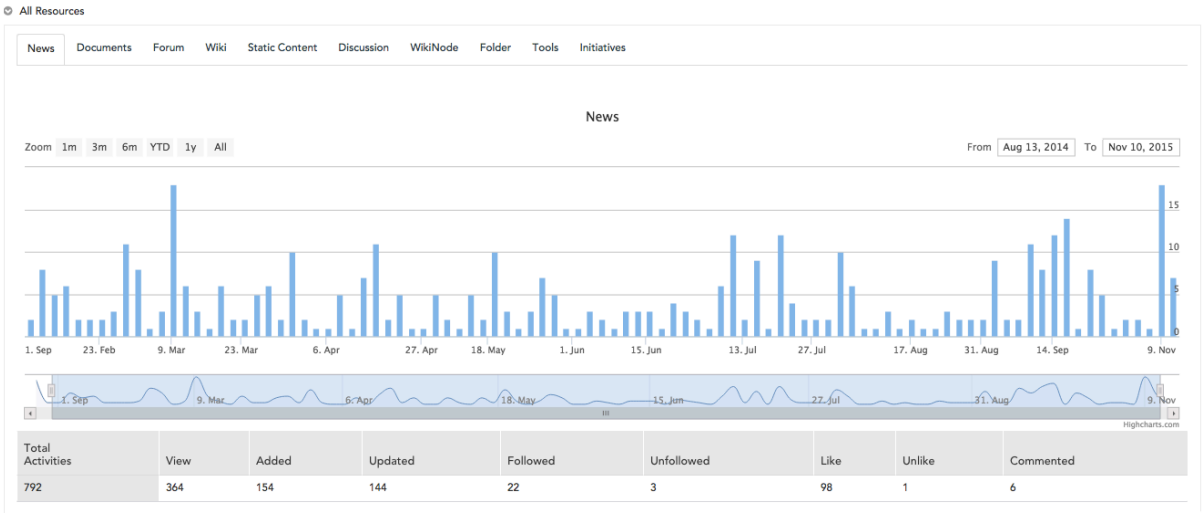


Figure 10 - Activities by resource type over time

ACDC Social analytics provides also detailed information on single resource, meaning that by clicking on the name/title of a resource, the manager will have a complete and unified view of all the activities done in relation to that resource. At this level the ACDC Social Analytics provides detailed information regarding the activities done and the users who acted on a resource.

These types of information can help understand the lifetime of the resource, shows in a unified manner the users who actively interact between them on a specific resource, helps retrieve the value of the resource itself and can lead the managers to take the correct decisions in order to better engage users into the services offered by ACDC Community Portal.

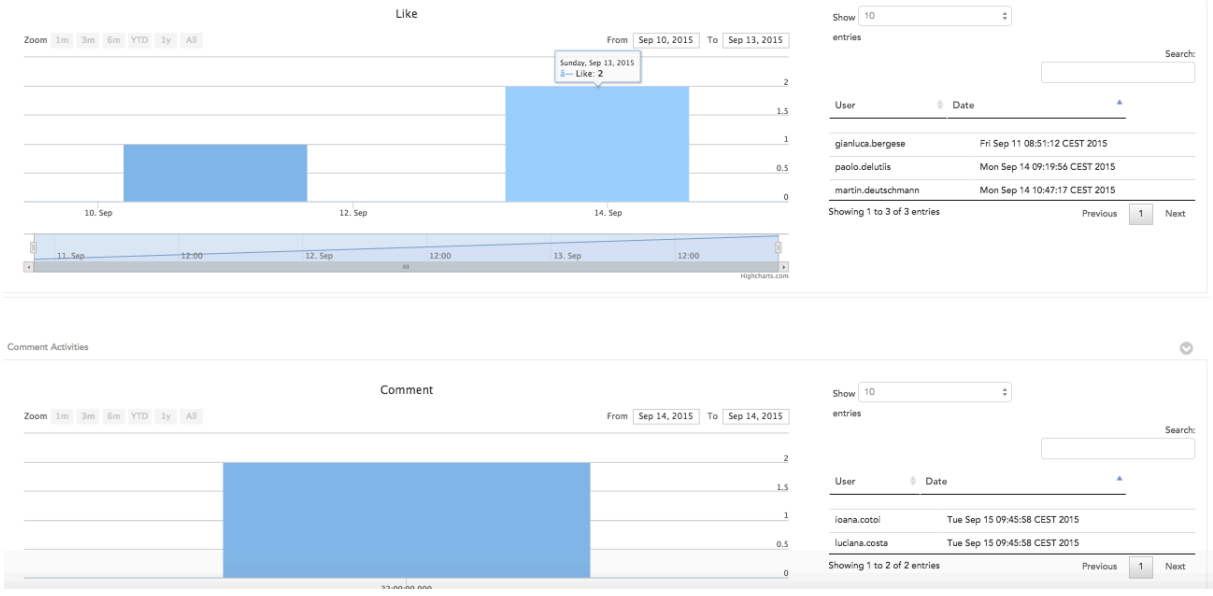


Figure 11 - resource activities, detailed view

5.2.1.3. Privacy concerns regarding the ACDC Social Analytics

Since the beginning, ACDC consortium has taken the protection of the personal data very seriously, committing to protect it.

ACDC Social Analytics focuses on the community by analysing the evolution of contents and the usage of the community portal in order to make sure users receive useful and needed services. Therefore, the purpose of the ACDC Social Analytics is to provide managers of the ACDC user community with insights on the community, to help them better understand the community needs and improve their decision making with respect to community administration.

ACDC Social Analytics processes data and information, by complying with the European data protection legal framework and principles and in a privacy considerate way.

Therefore:

1. ACDC Social Analytics does not track users over time and build profiles of their interests, characteristics, such as age and gender, and shopping activities;
2. ACDC managers do not display advertisements that reflect users' interests;
3. ACDC managers do not use the information retrieved from the ACDC Social Analytics against the interest of the users, but to improve the services and to satisfy the expectations of the users;
4. ACDC managers do not identify users beyond the ACDC Community Portal boundaries;
5. ACDC carries out the measurements on its own, and therefore does not use third-party web analytics tools;
6. Only the managers (EII and ECO) of the ACDC Community Portal have access to the personal data coming out from ACDC Social Analytics;
7. ACDC Managers do not make public or share with third-parties the personal data regarding the users coming out from ACDC Community Portal;
8. ACDC Social Analytics provides information, on both stakeholder and resource level, for up to one year.

5.2.2. Data Sharing

Accordingly to the purposes introduced in section 5.1, the data sharing part of the ACDC Analytics dashboard contains charts related to the DAM usage, that are briefly described in the subsections below.

As detailed in deliverable D6.2.1 [12], access to DAM functionalities is selectively enabled to some of the ACDC Community members. Access to DAM is a prerequisite for data sharing through the Central Clearing House for two reasons:

- the **API-Keys** released by DAM to community participants must be included in requests sent to the CCH service, otherwise requests to the CCH will not be authorized.
- the DAM allows community participants to setup **Sharing Policies**, thus enabling the forwarding of shared data to other community participants.

5.2.2.1. Overall DAM Usage

The access to DAM functionalities can be requested by each stakeholder of the community, and the request is approved (or denied) by the CCH Managers (a role in the community currently held by representatives of ECO in the community portal). Not all the organizations of the ACDC community requested access to DAM, because not all community members have data to share, or capabilities in place to process the data. Therefore, the statistics of DAM usage reports a number of organizations that is lower of the entire ACDC user community.

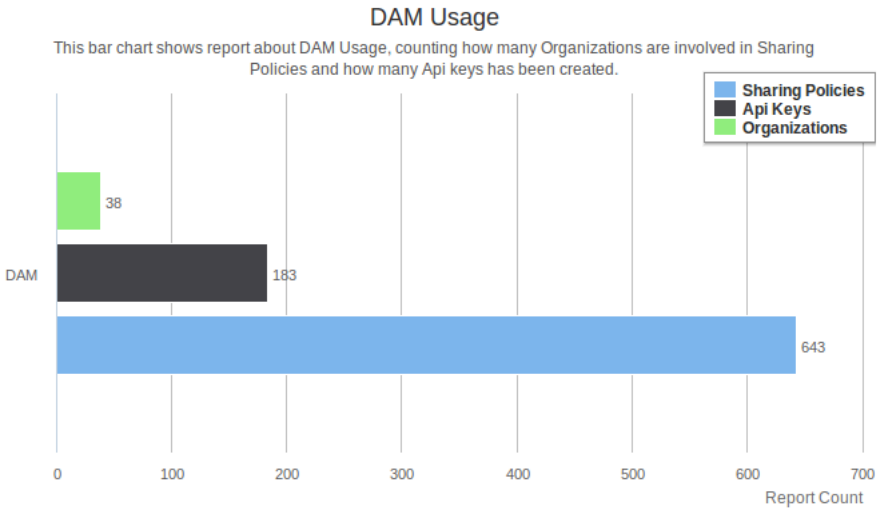


Figure 12 - DAM Usage Totals

The *DAM Usage* chart shows the overall usage of the DAM by the ACDC user community. It contains totals on the number of organizations using the DAM, the number of API Keys currently active and the number of sharing policies in place. This is a summary graph aiming at help community managers to **evaluate the whole community involvement** in the data sharing tasks.

5.2.2.2. API Keys – Read and Write

After acceptance by the CCH Managers, the stakeholder is able to access DAM functionalities, and create its own read and write API-Keys. These can be used to respectively retrieve data from the CCH (for Read API-keys) or provide data to the CCH (for Write API-Keys). API-Keys can be either in the "active" state, or in the "invalidated" one (because the API-key is expired, or because it has been revoked by user or by the CCH Manager).

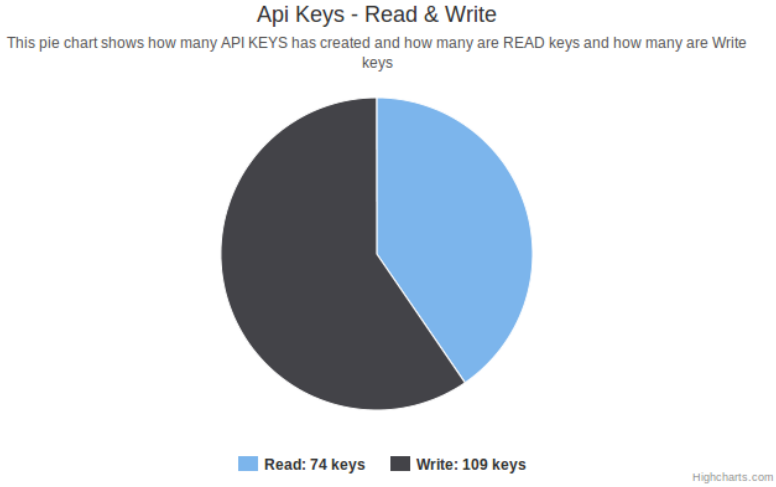


Figure 13 - DAM Read and Write keys

The *API-Keys - Read and Write* chart shows the totals for Read and Write API-keys currently active in DAM. By looking at this graph community managers can have a clue about **the balance between data**

providers (those supplying data to the CCH) **and data retrievers** (those processing the data supplied to the CCH). Again, this is a high level view and can be observed in conjunction with more detailed charts below (like those in sections 5.2.2.4 and 5.2.2.5).

5.2.2.3. CCH Groups and Organizations

Once a stakeholder is granted access to DAM, the stakeholder is placed in the "Unverified" CCH group. Each CCH Group is associated to a set of constraints in the usage of the CCH service and DAM application. For instance, stakeholders in the unverified group can create API-Keys of type Write only, thus being able to provide data to the CCH, but not to read other providers' data.

Given the multiple natures of some ACDC members (those for instance are both CERTs and ISPs), stakeholders can also be assigned to more than one CCH group (assignment can be changed by the CCH Managers through the DAM administrative interface). In case of assignment to multiple CCH groups, at API-Key creation the stakeholder can set the CCH group the new key belongs to (and limitations are then enforced on the subset of API-keys belonging to that group).

CCH Groups reflects the stakeholder cybersecurity positioning categorization, as reported in deliverable D6.1.1 [11]. The cybersecurity positioning for each community member is also present in the member's profile in the community portal (accessible by all community members in the Initiatives & Stakeholders section).

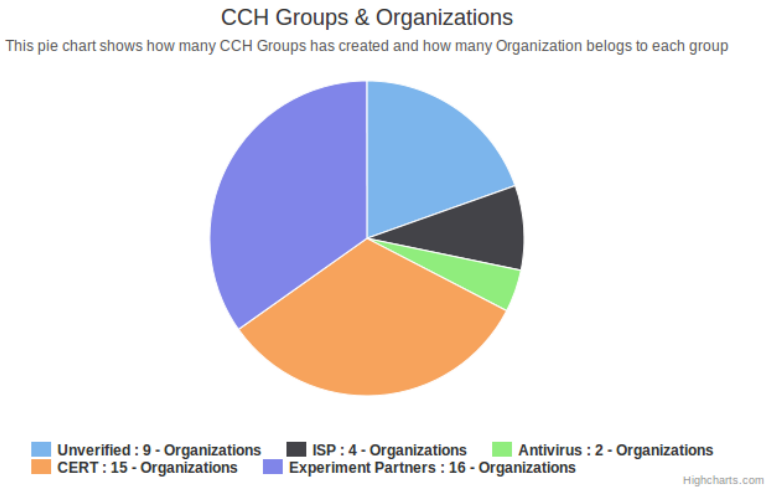


Figure 14 - DAM Stakeholders by group

The CCH *Groups & Organizations* chart shows the number of stakeholders belonging to each CCH group. As stated above, each stakeholder may belong to more than a CCH group, for this reason the sum of organizations over all the categories is higher of the total number of stakeholders reported in the DAM Usage chart. Nevertheless, this graph helps community managers to assess the involvement of the different kind of stakeholders to the data sharing task .This is important either to **correct lack of participation** from members that are already in the community, or to **properly plan involvement campaigns**, that could aim at involving new stakeholders of a given kind to improve the data sharing and analysis tasks.

5.2.2.4. CCH Groups and API Keys

Every stakeholder with access to DAM can create multiple API-Keys, within the limitations of each CCH Group (as from previous section), and use them for different purposes. For instance, different write keys can be used to partition the data provided by different sensors (by using one API-Key for each sensor), thus allowing the stakeholder to setup separate sharing policies for data coming from different sensors. Similarly, multiple read API-Keys can be used to partition the data that are received from other community participants, thus easing the logic of processing services.

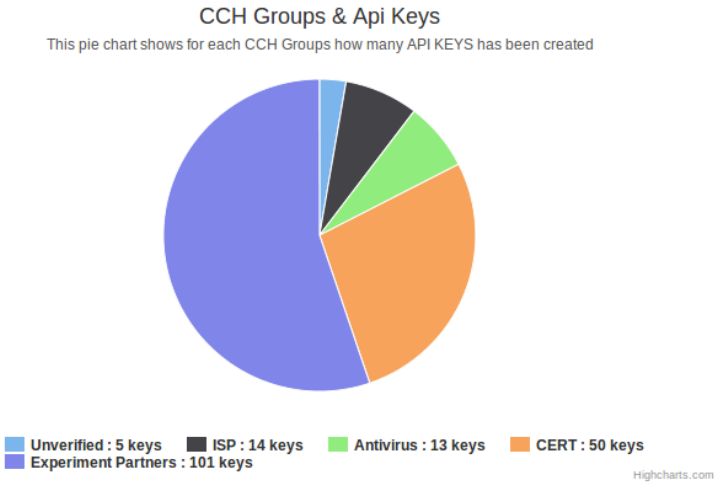


Figure 15 - DAM API Keys by group

The *CCH Groups & Api Keys* chart shows the distribution of keys over the different CCH groups. The usage of keys within each group, if used in conjunction with previous graphs (see above sections), could **highlight anomalies** in the usage of DAM, e.g. small groups with large number of API-Keys (or the contrary) could be a symptom of a wrong usage, that may trigger further investigation by community managers and CCH Managers.

5.2.2.5. Sharing Policies for each group

Once APi-Keys have been created by each users, they can be used to request (and establish) sharing policies with other community members. The default routing of information in the CCH forward reporting (e.g. about cyber-attacks) only to the organizations that are responsible of ASN and IP addresses indicated in the report itself. This scheme works well for CERTS and ISP, but limits the contribution to the data processing by other organizations like, for instance, antivirus companies, that are usually not responsible for any ASN and IP range, but needs data to update their antivirus products.

To address this limitation, stakeholders of the ACDC community can request to share data with other participants. The request is bound to a couple of API-keys (one read and one write API-key), and can either originate from the data provider or from the data retriever. The other party involved in the request is notified and can either accept or deny the data exchange. If the data exchange request is accepted, a new Sharing Policy is created and notified to the CCH. This has the effect of forwarding all the reports submitted by using the write API-key to the channel associated to the read API-key.

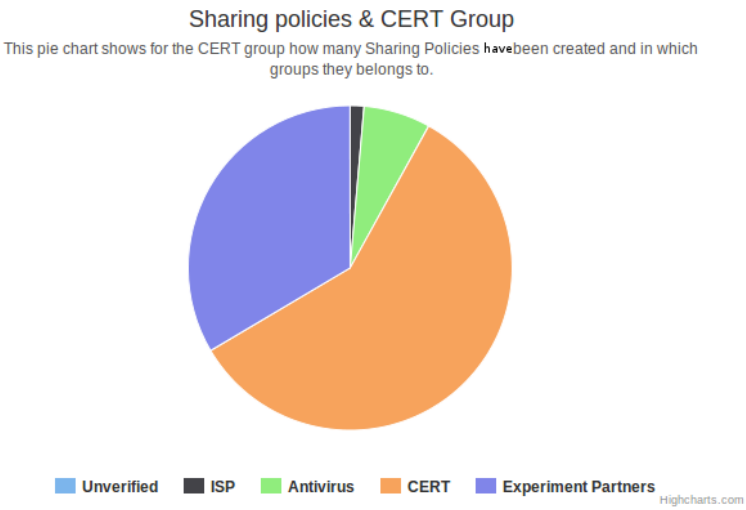


Figure 16 - DAM sharing policies by group (sample)

The set of graphs *Sharing Policies by "CCH Group"* shows, for each kind of stakeholders, the distribution of the Sharing Policies in place with respect to all the CCH groups. The sample in the figure above shows, for instance, the distribution of Sharing Policies for the CERT group. More than half of the sharing policies are set within the CERT group itself (thus between CERTS) and a large part points toward the Experiment Partner group (that has been created to run ACDC experiments). The added value of this representation is to highlight the **sharing attitude** of stakeholders in each CCH group, that can be used to assess the most common flow of information among different kind of stakeholders.

5.3. ACDC Social dashboard implementation

The implementation of the ACDC Social Analytics is based on the information stored in the community portal database. In particular, two sources of information have been used to generate populate graphs available the analysis application:

Portal Audit Log – this part of the portal database contains all the events related to the user activities on contents. This includes:

- authentication actions (login, logout)
- creation and modification of resources (news, message in the forum, initiatives, etc.)
- social actions on resources (like, unlike, follow, comments, etc.).

The audit log information has then been used to populate information for the *User Content* part of the ACDC Social Analytics application. The content of the audit log is periodically checked by a scheduler and the content older than the defined timeframe is removed. The timeframe (set by default to one month) can be extended up to one year, thus also implementing a retention limit for the information related to user activities. Of course the actions of users on resources are kept longer than that, only their traces in the audit log are removed, and will disappear from the ACDC Social Analytics views.

DAM database – the DAM database is the part of the community portal database storing information about:

- associations between stakeholders and CCH Groups
- API-Keys created by each user of the ACDC community (of course actual API-Key values are not stored, they're, instead, retrieved from the CCH upon user's request)
- Sharing policy information (including pending requests for sharing, as well as sharing policies in place) .

The actual content of the DAM database is used to generate graphs for the *Data Sharing* part of the ACDC Social Analytics application. This content is only deleted upon user request, as API-Keys and Sharing Policies are essential for the operational of the CCH service and its clients.

5.4. Monitoring the ACDC User Community

5.4.1. Analysing the ACDC Social Interaction

ACDC Analytics reporting dashboard is divided into 3 sections: reports, stakeholders and resources. These three dimensions offer an extensive set of dashboards that allow the manager for a deeper analysis of the engagement of users, the platform capabilities, as well as administration activities (total resources published, deleted, modified etc.).
ACDC Analytics is available for user engagement and satisfaction analysis, content optimization, campaign and post-campaign analysis. The following sub-sections will provide concrete insides regarding the evolution of the Community Portal during the last year and its usage.

a) The reports

The "Report" approach (see figure below) is an easy and fast way to monitor the evolution of the Community Portal, in terms of users, organisations, initiatives and tolls and services. This approach was not described in the previous sections, as in se per se it's not an analytic tool, but more a statistical one, which provides the first insides regarding the Community Portal and its evolution. Together with the other tools developed, the report approach helps the manager to have a complete picture of the portal progress.
Therefore, in the figure below it can be noticed that after a year from the official launch of the portal (September 2014), 165 users and 179 organisations registered to the portal, 61 tools & services were added by users and 35 initiatives in the cyber field were indicated.

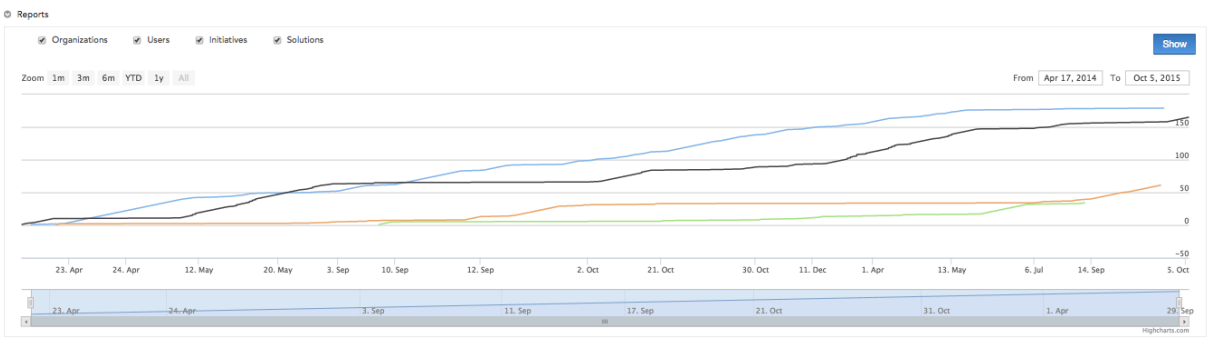


Figure 17 – Results regarding the Reports

This approach was very useful, especially during the evaluation of the impact of several events (launch, dissemination activities, campaign) done from September 2014 until September 2015. Based on the figure, it can be noticed that on May 2014 the number of users increased to around 60, which coincides to the internal launch of the portal; after that, by middle of October 2014, the numbers of users increased to around 90; at the end of February the number of users increased to 110 due to the events where ACDC was represented and disseminated; the Dedicated Support Campaign, which

took place from March to May 2015, helped increase users by 30% (reaching 150 users at the beginning of June).

b) The stakeholders activities

The stakeholders' analysis approach focuses on the relationship between organisations and users and facilitate the community managers to (i) analyse the level of animation based on the stakeholders activities, (ii) understand the level of interaction of each stakeholder, (iii) observe the active and passive stakeholders in a given timeframe, (iv)identify the negative (or positive) trend of a stakeholder during time.

Stakeholders Activities

Show 25 entries

Search:

Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total activities
Engineering Ingegneria Informatica	141	277	11	27	1	58	1	7	523
Atos	109	35	0	0	0	4	0	1	149
DFN-CERT TEST	80	25	0	0	0	0	0	0	105
Telecom Italia Information Technology	14	35	0	3	0	26	1	7	86
Incibe	63	15	0	0	0	0	0	1	79
XLAB	53	12	0	6	0	6	0	1	78
CARNet	5	43	0	2	1	7	0	0	58
University of Luxembourg	54	0	0	0	0	0	0	0	54
eco e.V.	24	8	0	0	0	3	0	1	36
Telefónica I+D	20	0	0	0	0	2	0	0	22
TECHNIKON Forschungs- und Planungsgesellschaft mbH	0	0	0	0	0	13	0	0	13
G Data	7	1	0	0	0	1	0	2	11
Montimage	1	0	0	1	0	4	1	2	9
LSEC - Leaders In Security	4	1	0	0	0	0	0	1	6
synaix Gesellschaft für angewandte Informations-Technologien mbH	4	0	0	0	0	0	0	0	4
Katholieke Universiteit Leuven (ICRI-CIR)	0	0	0	1	1	2	0	0	4

Figure 18 – Results regarding the Stakeholders activities

During the last year the "Stakeholders activities" approach was used to define and monitor a campaign, to observe the positive (add, like, comment, follow) and negative actions (delete, unfollow, unlike), identify the stakeholders who are more oriented to provide content, compared to those who are more oriented to socially interact around it, and adapt the campaign based on that).
By analysing the data it came out that the community-building process is stimulated by the existence of a common activity on which to work together, and for which there is a clear added value in sharing results, and comments, about the work. This is also reflected by the fact that the most active stakeholders in the Community Portal were those also involved in the data sharing and in the ACDC experiments . The Community Portal was the main tool used for the exchange of information, data and communication for these activities, reason why the partners involved in those tasks have a higher level of engagement.
Moreover, we also understood that the involvement and the participation of a user into the community can be determine by several aspects, like:

- their role in the organisation – not all users registered to the portal have an operational role inside their organisation (i.e. in Atos, Rodrigo Dias is the Head of Lab and does not cover an operational role in the organisation, reason why his engagement is 0)

Stakeholder Analytics: Atos

All Counters Activities

Show 10 entries

Search:

Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total Activities
beatriz.gallego	103	31	0	0	0	0	0	0	134
elsa.prieto	6	4	0	0	0	2	0	1	13
rodrigo.diaz	0	0	0	0	0	0	0	0	0
carlos.arce	0	0	0	0	0	0	0	0	0
susana.gonzalez.zarzosa	0	0	0	0	0	0	0	0	0

Showing 1 to 5 of 5 entries

Previous 1 Next

Figure 19 – Stakeholder Analytics ATOS

- their role in the project – not all users registered to the portal have an active role in the project (i.e. in ENG, Mario Rodrigues and Angelo Marguglio are not directly involved into the ACDC project – not any more, or had a limited involvement into the project, reason why they have a low level of engagement)

Stakeholder Analytics: Engineering Ingegneria Informatica

All Counters Activities

Show 10 entries

Search:

Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total Activities
paolo.roccetti	98	197	6	2	0	4	1	1	309
vincenzo.napolitano	36	48	0	12	0	15	0	1	112
ioana.cotoi	6	27	5	10	1	35	0	2	86
mario.rodriguez	0	4	0	0	0	0	0	2	6
demo.user	1	0	0	0	0	0	0	1	2
angelo.marguglio	0	1	0	0	0	0	0	0	1

Figure 20 – Stakeholder Analytics ENG

- the background of the users and the profile of the organisations can be very important, as the purpose of the portal is very concrete and technical in the same time, for some users is much more difficult to interact (i.e. KU Leuven, which cover the legal aspects in the project compared to Telecom Italia, which have the expertise and knowledge to share info and data regarding botnets)

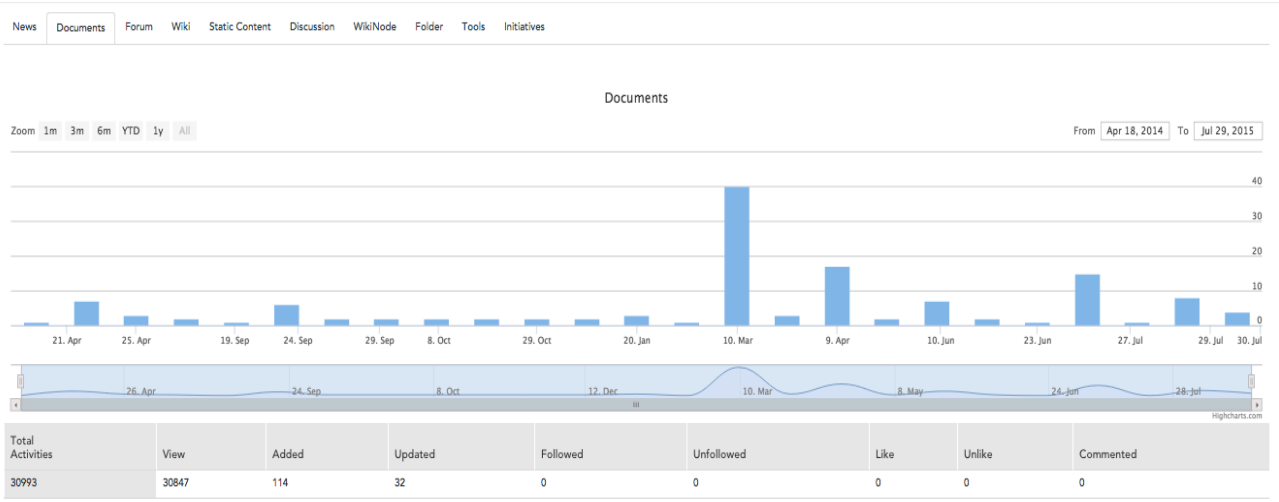
Name	Added	Edited	Deleted	Followed	Unfollowed	Like	Unlike	Comments	Total activities
Katholieke Universiteit Leuven (ICRI-CIR)	0	0	0	1	1	2	0	0	4
Telecom Italia Information Technology	14	35	0	1	0	25	1	6	82

Figure 21 – Comparison on the stakeholders results

c) Resources

The “resources” approach focuses on analysing the resources in each feature, such as News, Initiatives and Stakeholders, Tools and Services, Documents, Experiments, Forum and DAM. Through this approach the manager can understand the usefulness of the features, the resources that produce more rumours and understand the behaviour of the users based on the types of interaction they prefer during the time.

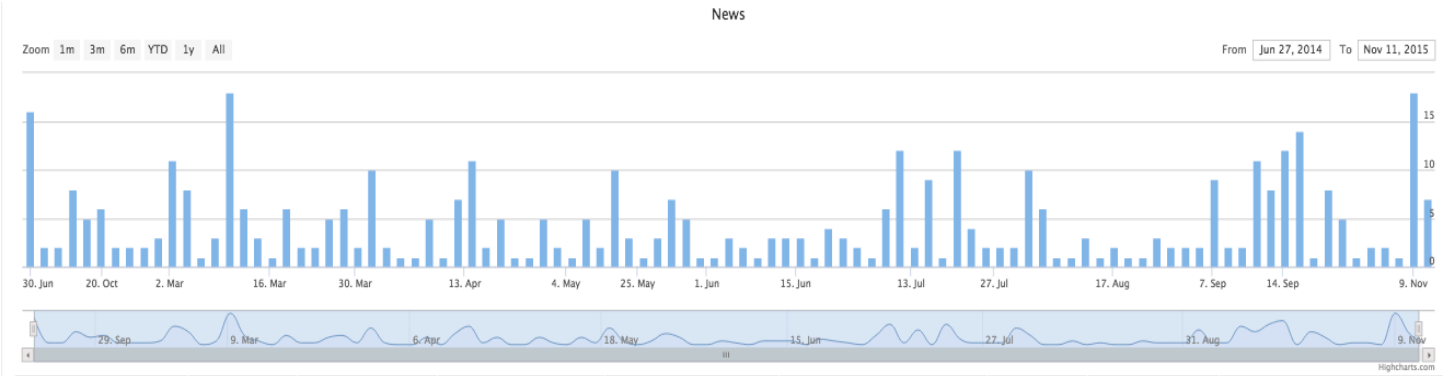
The analysis shows that the feature used the most is “News”, followed by “Forum” and “Tools”. Of course, there is a strong difference between the features mentioned before, and the “initiatives”, “Wikis” and “Documents”, as the last ones are static and there is no interaction around the content.



Total Activities	View	Added	Updated	Followed	Unfollowed	Like	Unlike	Commented
30993	30847	114	32	0	0	0	0	0

Figure 22 – Results on «Documents»

The only indicator that can reveal the usability of the features is the number of visualisation of the content, which for example, in “Documents” there were added 114 files that had 30150 views. It is also important to mention that ACDC partners used another repository tool for sharing files of interest of the project, while ACDC Community Portal was used only to share files of public interest.



Total Activities	View	Added	Updated	Followed	Unfollowed	Like	Unlike	Commented
792	364	154	144	22	3	98	1	6

Show 10 entries

Search:

Name	Create Date	View	Updated	Followed	Unfollowed	Like	Unlike	Commented	Total Activities
botnet of infected PoS	2014-06-27 12:03:49.0	27	1	1	0	1	1	0	32
New way of Dridex banking malware infection	2015-03-30 12:18:35.0	11	1	1	0	1	0	0	15
Spam Campaign hits German Telekom - do you see something similar in your country?	2015-08-20 10:27:59.0	11	1	0	0	2	0	0	15
Phishing Education Can Save Nearly $4m Annually	2015-08-27 15:41:22.0	7	1	1	1	2	0	0	13
Google, Yahoo, Facebook Collaborate to Blacklist Bad Bots	2015-09-10 18:02:49.0	5	1	1	0	3	0	2	13
What is the Data Protection Authority?	2014-10-16 12:57:42.0	9	1	0	0	0	0	0	11

Figure 23 – Results on «News»

This section gathers all the latest steps taken in ACDC, new features of the community portal, latest events, latest news about cyber security, availability of new solutions, etc. In the above figure it can be noticed a high level of interaction especially on May, July and September 2015, which corresponds to the post-campaign period.

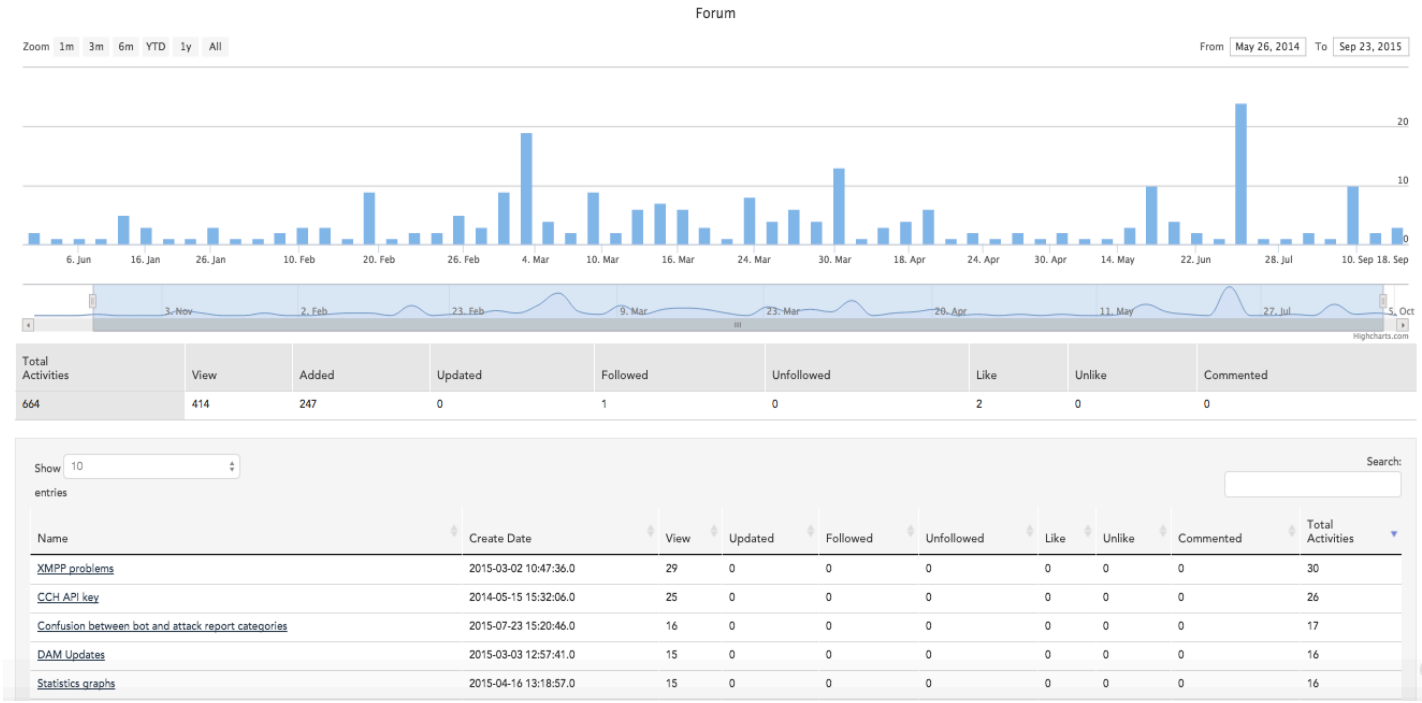


Figure 24 – Results on «Forum»

"Forum" is the place collecting all the discussions among community members. The stakeholder uses this application to see all the comments, messages, threads and related posts published in the portal. The period with the highest interaction on "Forum" is from February 2015 to April 2015, mostly on data sharing and experiments topics. This is the period were some of the ACDC Experiments were run, and in fact discussions added to the Forum in that period are related to the execution of the ACDC experiments.

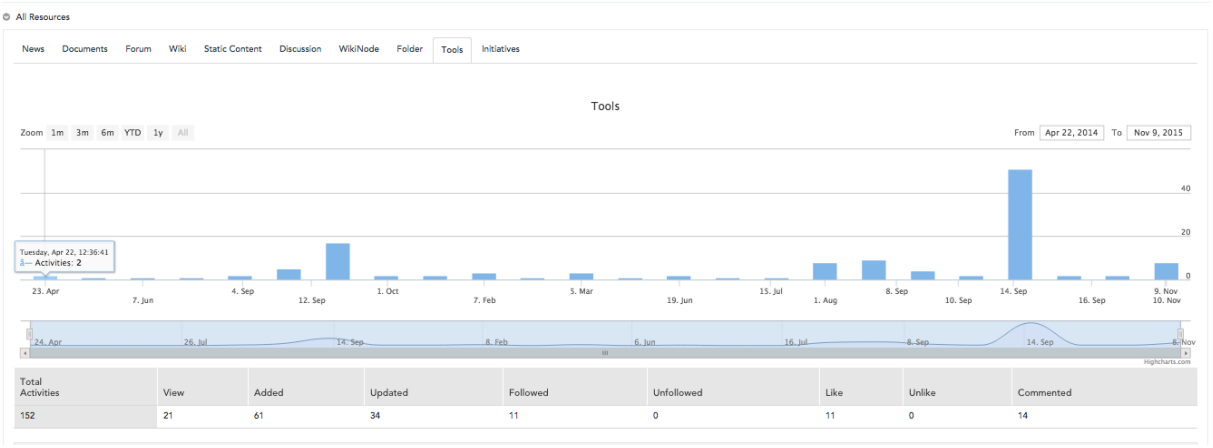


Figure 25 – Results on «Tools&Services»

Through "Tools & Services", ACDC brings a list of available tools and services that can be implemented to protect against cyber attacks; also users can add new tools and services provided or used by them. This feature was used more by ACDC consortium which listed the available tools.

5.4.2. Analysing the ACDC Data Sharing

As introduced in section 5.2.2, the ACDC Analytics also includes the reporting functionalities related to the Data Access Manager (DAM) usage. DAM is used by the ACDC community to setup the access keys and sharing channels among data sharing participants. Monitoring keys and sharing policies during the lifetime of the ACDC project provides an insight about the level, and variety, of interconnections among the ACDC members at the data sharing level.

The following figure shows the DAM overall numbers at the beginning of September 2015 (end of the ACDC project.

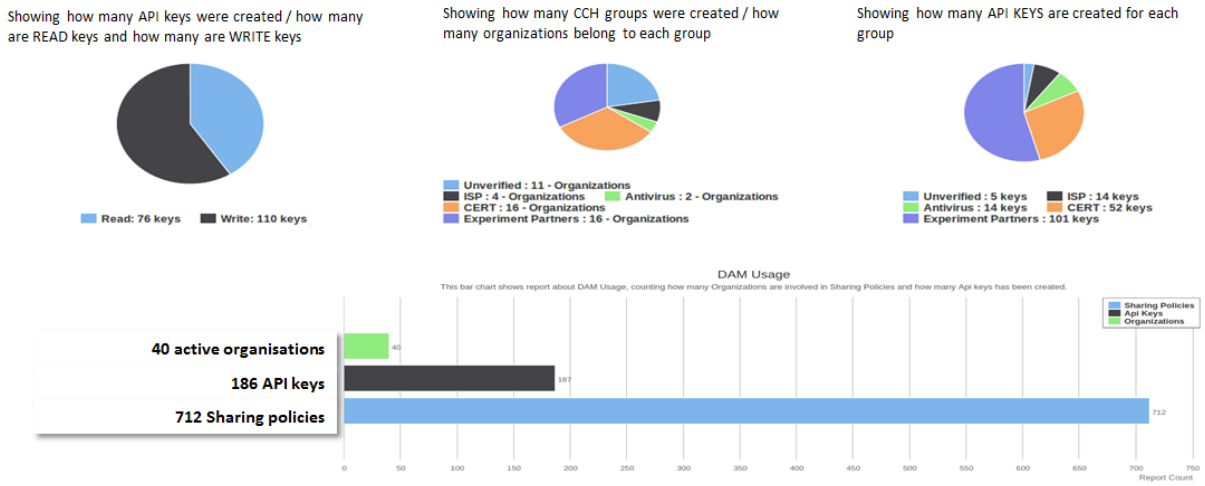


Figure 26 - DAM Statistics overview

The Figure shows (in the pie graph on the top left) a higher portion of keys used to report data to the CCH (110 keys) compared to the number of keys used to read data to the CCH (76 keys). This is in line with the approach proposed by the ACDC project, seeing a potentially large set of reporting organizations, compared to the number of those able to analyse the data and correlate them.

Also, the picture shows (in the pie graph at the top centre) the number of organizations in each CCH group. As from the diagram the most represented groups in data sharing are CERTs (16 Organizations). The *Experiment Partners* group were created to group all ACDC partners sharing data. The *Unverified* group has also a high percentage, it is meant to contain the organisations that are not yet trusted enough to be moved to a different group (organisations that are in the Unverified Group cannot create read keys, thus they cannot access data, only provide them to the CCH). This second pie chart highlights the need for a more efficient mechanism to move organisations from the Unverified Group to the other "more trusted" groups, as well as to enlarge the set of groups initially defined (that includes CERTS, ISP and Antivirus) to better detail the data sharing participants.

The third pie chart on the top right shows the number of keys created by each group. Reasonably, the vast majority of keys were created by the Experiment Partners group (to run ACDC experiments), followed by the CERT group (52 keys) and ISP & Antivirus groups (14 keys each). The number of keys created are in line with the indications given for the ACDC Experiments (each different data format should be reported by using a different key, to ease the data parsing on the recipient's systems).

Finally, the bar chart at the bottom of the figure shows the totals for Organizations sharing data, API Keys and Sharing Policies created. The number of organisation differs from the total in the CCH Groups pie chart, as every organization may belong to more than one group (except the ones in the *Unverified* group).

The Data Sharing also reports about the number of Sharing policies currently active among the different CCH groups. The figure below reports the Sharing Policy distribution at the beginning of September 2015.

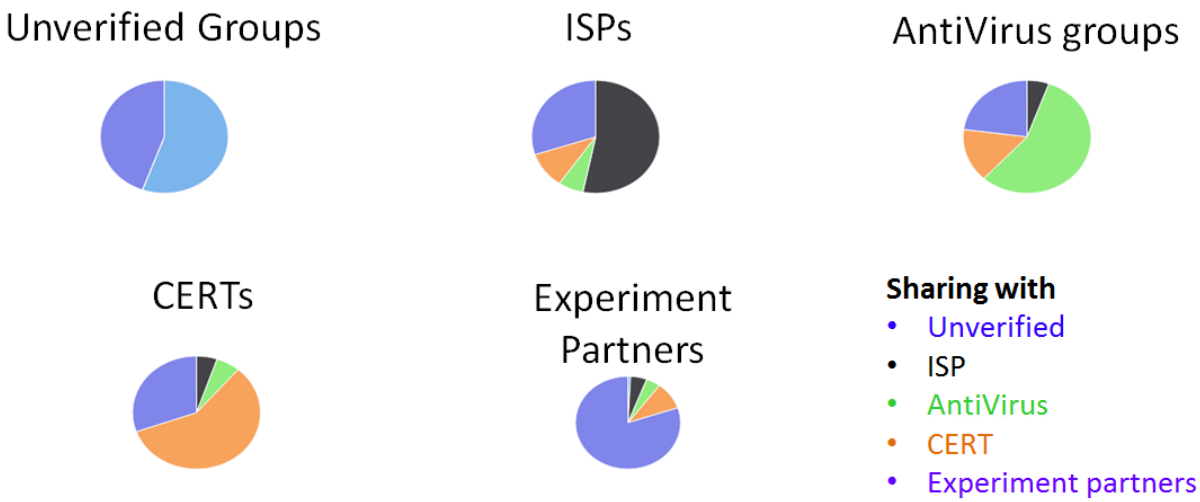


Figure 27 - DAM Sharing Policies distribution

Each pie chart shows the distribution of sharing policies (those active at the time of the image were taken) for each CCH group. For instance, the ISPs chart on the top centre shows the distribution of sharing policies from the ISP group to its members (black slice) and to members of the other CCH groups (orange for CERTs, Green for Antivirus, etc.).

As shown by all the pie charts, there is a tendency to "intra-group" sharing (meaning CERTs mainly shares with other CERTs, ISPs with ISPs, etc.) Therefore it may be worth to investigate further on this tendency, with the aim of stimulate the setup of sharing channels among different CCH groups, that could lead to dissemination of data supplied to the CCH to a wider set of stakeholders.

5.4.3. Analysing the ACDC Botnet Metrics

The Botnet Metrics section in the ACDC Community Portal provides community members with access to statistical information about CCH reporting. The information model is based on the Botnet Metrics activity carried out in WP4 of ACDC, which identified a set of metrics to "measure" botnets.

The development of the Botnet Metrics application has been carried out in WP4, and is reported in D4.4 – Publicly accessible database of botnet metrics [14]. Nevertheless, being the botnet metrics information based on the reporting to the CCH, it is worth to provide some example analysis here. These examples are provided with the double aim of (i) showing as the information generated by the Botnet Metrics section can help in assessing the reporting capabilities of the ACDC community and (ii) helping community members to better understand implications behind the figures provided by the application.

The Botnet Metrics application is able to generate a number of different graphs, i.e. "measures", referring to three different group of metrics: Data Quality, Botnet Impact and Cybersec Events.

- **Data Quality** metrics: assess the quality of the data submitted to the CCH in terms of gaps and anomalies, taking into account that it could distort the statistical stability of the data. This category of metrics can be used to comparing technologies (i.e. network sensors, malware analysis tools, spam-traps, correlation services, etc.) based on the information reported to the CCH, both on qualitative and quantitative terms.
- **Botnet Impact** metrics: assess the impact of the botnet activity in terms of distribution, by comparing incidents related to Bots (Unique IP-based, Proxy-based, RDNS-based) per ASN, per Country, per ISP-subscriber.
- **Cybersec Events** metrics: this type of metrics group those that focus on evaluating volume and quality of data reported to the CCH in the context of specific pilot experiments, such as those conducted in WP3, or certain types of cybersecurity related events such as DDoS attacks, Malicious URIs detected or Malware samples analysed.

In the following, two examples will be presented: a "Gaps in Time" comparison between two keys (from the Data Quality group above) and a "Unique IP per Country" comparison (from the Botnet Impact group above).

Gaps in Time

In the Gaps in time example, the user selects a period of time and two API-Keys among those that have been used to report data to the CCH in the selected period (that are shown to the user in a multi-selection menu). The application then generates a graph showing the number of reports by each of the selected key on a daily scale. The generated graph is shown in Figure 28.

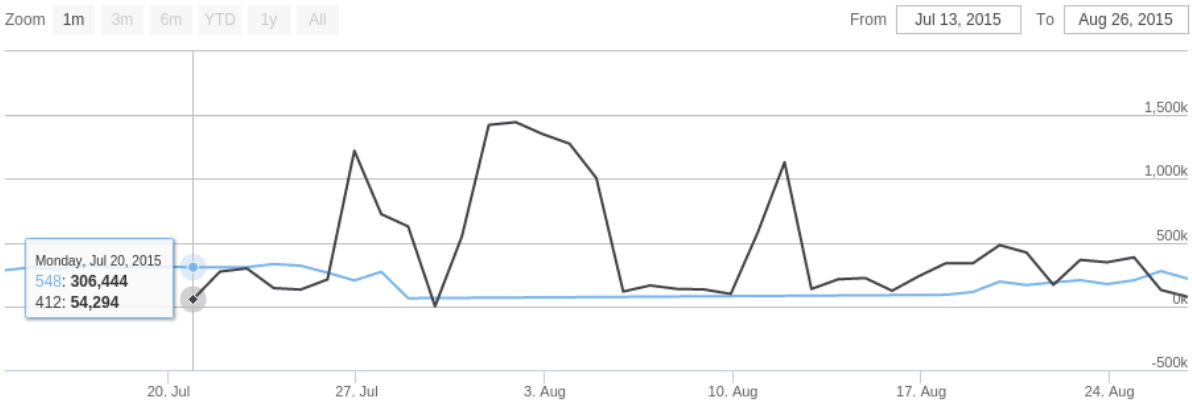


Figure 28 - Gaps in time graph

The graph clearly highlights differences in the reporting behaviour of the selected keys. Key 412 (black line) shows huge variations in the reporting frequency, with peaks over 100.000 reports per day. On the other hand, key 548 (blue line) shows a more regular reporting, with average values around 250.000 reports per day. This difference, immediately evident, could help to monitor network sensors that are using these API-Keys, and to discover anomalies and misbehaviours of those sensors. In fact, after investigating the above keys, the behaviour turned out as normal, because key 412 is used by a sensor that reports about DDoS attacks (thus its irregular reporting behaviour is justified), while key 548 is used by a honeypot system (thus having a more regular reporting frequency). The comparison can be done among all the keys that reported data in a given period of time (chosen by

the user at the beginning of the search). This gives sensor administrators and analysts an easy way to investigate the *quality* of data reported to the CCH.

Unique IP per Country

In the Unique IP per Country example, the user again selects a period of time and a number of countries of interest, among those having at least one report in the timeframe selected). The application then generates a graph showing the number of reports, in the reporting period, for each of the countries selected.

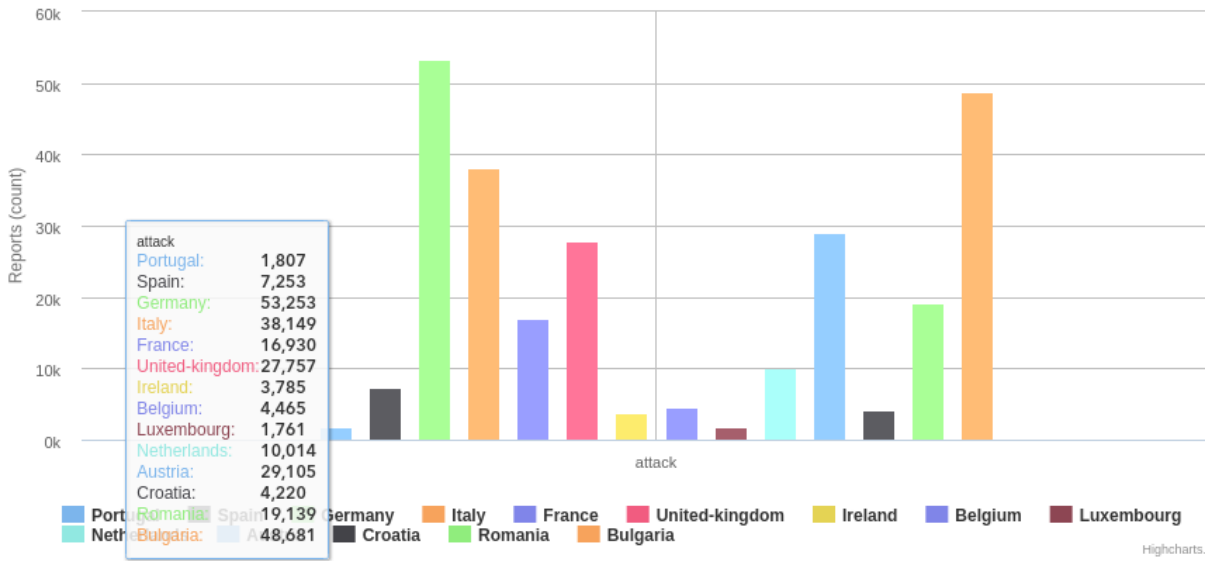


Figure 29 - Unique IP per Country example

In the figure above, a sub-set of European Countries were selected, and the total number of reports (in the attack category of reporting) is shown for each one in a bar chart. One of the interesting aspects of this picture is that, as most of the attacks cross country boundaries, the statistical value of the information is not influenced that much by the number of sensors deployed in a given country (as each sensor also reports about suspicious IP in other countries). The information could then be used to estimate where the majority of attacks originates (in this timeframe: Germany, Bulgaria and Italy), and thus to identify the countries where a "cleaning" campaign could have chances to be more effective in reducing the botnet phenomena.

6. Conclusions

In order to support the ACDC user community in their interaction and participation mode, a set of social analytics tools were developed and integrated into the social platform to ease the analysis of the ACDC community in terms of growth, evolution and behaviour. Therefore, ACDC Analytics was developed so that it can provide figures about the community participants, the contents actually stored in the community portal, the social interaction, the data interaction using the CP / the CCH, statistics about the data in the CCH etc.

The community portal was deployed in April 2014, and is currently used by ACDC partners and community members. It is maintained, after the end of the project, and continues to evolve as one of the key pillars for the future of ACDC. In this context, the release of ACDC Analytics was postponed to early 2015, to give the consortium and the community a timeframe for the adoption of the portal in ACDC activities.

ACDC Analytics was and is used to support managers of the ACDC user community and to provide them with insights on the community, to better understand the community needs and improve their decision making with respect to the community administration.

The objective regarding the ACDC user community, the Community Portal and the Social Analytic tool, created and developed during the 30 months of the project, is to migrate to a self-sustainable environment to support the **single-access point, cyber eco-system** sustaining and fostering concrete **collaborative actions between stakeholders who will then** increase their level of protection against cyber-disruptions, being part of a **European network** of individuals and organisations, called CyberConnector.

The collaborative actions proposed by CyberConnector are based on **decisions, solutions and knowledge** elaborated from a collaborative pooling of information.

The strategy used for CyberConnector is based on **building trust in a bottom-up approach** that allows smaller groups to work together, in parallel, on very focused activities, in their own workspace, and then expand out to a cyber-security environment once they are in the habit of working together.

The CyberConnector environment will operate at European level, to exploit the migration of the knowledge created in different communities through European projects – namely CYSPA focused on risk assessment, COURAGE and CAPITAL focused on cyber-security and cyber-crime, DOGANA focused on Social Vulnerability risk assessment and ACDC focused on botnets.

By combining the ACDC values with CyberConnector, therefore with a wider offer in terms of knowledge, and by designing new forms of collaboration in a larger community, users will be provided with a increasingly comprehensive environment. Finally, providing users with a single access point to an overall cyber security environment will further improve their customer experience by enlarging the scope, value and coverage of content included and accessible through this environment and will therefore generate an increase in terms of customer engagement and satisfaction.

Therefore, **ACDC Analytics will be integrated into CyberConnector,** to ease the analysis of different communities in **terms of growth, evolution and behaviour,** in order to improve **the services** and the **users' satisfaction**.

7. References

[1] Social Analytics Wikipedia, http://en.wikipedia.org/wiki/Social_analytics
[2] Measure Social Activites in Liferay, http://www.liferay.com/it/documentation/liferay-portal/6.1/user-guide/-/ai/measuring-social-activity
[3] Measure Social Interactions with Google Analytics, https://developers.google.com/analytics/devguides/platform/social-interactions
[4] Google Analytics Reporting API, https://developers.google.com/analytics/devguides/reporting/
[5] Klout, http://klout.com
[6] Kred, http://kred.com/
[7] PeerIndex, http://peerindex.com/
[8] Google Analytics Goals, https://support.google.com/analytics/answer/1012040?hl=en&ref_topic=1007030
[9] CrazyEgg, http://www.crazyegg.com/
[10] ACDC DoW, 2012
[11] ACDC D6.1.1 - User Profiles and Categorization, 2013
[12] ACDC D6.2.1 - D6.2.1 - ACDC Social Platform Deployed, 2014
[13] ACDC D6.3.1 – Involvement model for users in ACDC, 2013
[14] ACDC D4.4 – Publicly Accessible Database of Botnet Metrics, 2015

[15] Awareness Inc. (2012). *Actionable Social Analytics: From Social Media Metrics to Business Insights*. Available on: http://igo2group.com/wp-content/uploads/2012/10/Actionable-Social-Analytics.pdf (accessed on 01/30/2014)

[16] Barber, I. (2010). *Opinion mining -Sentiment mining*. Available on: http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining (accessed on 01/29/2014)

[17] Few, S. (2007). Dashboard Confusion Revisited. Available on: http://www.perceptualedge.com/articles/03-22-07.pdf (accessed on 01/30/2014)

[18] Gartner. (2013). *Who's Who in Social Analytics*. Available on: https://www.gartner.com/doc/2597915 (accessed on 01/30/2014)

[19] Melville, P., Sindhwani, V., and Lawrence, R. D. (2009). *Social Media Analytics: Channeling the Power of the Blogosphere for Marketing Insight*. Available on: http://people.cs.uchicago.edu/~vikass/sma-win09.pdf (accessed on 01/30/2014)

[20] Noble, S. (2012). Google Analytics Dashboard to Monitor Brand Engagement. Available on: http://www.stateofdigital.com/google-analytics-monitor-brand-engagement/ (accessed on 01/29/2014)

[21] Olyslager, P. (2012). *Heatmap, Hot or Not?*. Available on: http://www.paulolyslager.com/heatmap-hot-or-not/ (accessed on 01/29/2014)

[22] Rouse, M. (2012). *Social media analytics*. Available on: http://searchbusinessanalytics.techtarget.com/definition/social-media-analytics (accessed on 01/29/2014)

[23] Stefanidis, A, Crooks, A., Radzikowski, J., Croitoru, A., and Rice, M. (2012). Social Media and the Emergence of Open-Source Geospatial Intelligence. Available on: http://css.gmu.edu/andrew/research/USGIFmonograph.pdf (accessed on 01/31/2014)

[24]Waisberg, D. (2013). *Google Analytics Dashboards: A Step-By-Step Guide*. Available on: http://online-behavior.com/analytics/dashboards (accessed on 01/29/2014)

[25]Walker, L. (2013). *Social Metrics: What Does It Mean?*. Available on: http://personalweb.about.com/od/socialmediaresearch/a/Social-Metrics.htm (accessed on 01/30/2014)

[26]Walker, L. (2013). *Social Analytics: Making Sense of Social Media Conversations*. Available on: http://personalweb.about.com/od/social-marketing/a/Social-Analytics.htm (accessed on 01/29/2014)

[27]Wang, Z. (2012). *Social Media Data Analysis: A Dynamic Perspective*. Available on: http://ftp.rta.nato.int (accessed on 01/31/2014)

[28]Highcharts library, http://www.highcharts.com/