REVEAL

FP7-610928

REVEALing hidden concepts in Social Media

Deliverable D9.4.1

Public project reports

Editor(s):	Christos Georgousopoulos
Responsible Partner:	INTRASOFT International S.A.
Status-Version:	Final – v1.0
Date:	17/12/2014
EC Distribution:	Public

Project Number:	FP7-610928
Project Title:	REVEAL

Title of Deliverable:	Public project reports
Date of Delivery to the EC:	17/12/2014

Workpackage responsible for the Deliverable:	WP9 – Project Management and Assessment
Editor(s):	Christos Georgousopoulos
Contributor(s):	ALL partners
Reviewer(s):	ALL partners
Approved by:	All Partners

Abstract:	This document summarizes the activities that took place within the REVEAL project in the first year [November 2013 – October 2014]. Achievements and work progress are described for all relevant work packages.
Keyword List:	12-month achievements report, November 2013 - October 2014

DOCUMENT DESCRIPTION

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	*Modified by*
v1.0	17/12/2014	preparation of D9.4.1 based on the periodic activity management report of year one	INTRASOFT

CONTENTS

LIST OF TABLES

1 Publishable summary report

In the beginning, the primary communication channel was the so-called 'word of mouth'. Societies were mostly rather small, and it was possible to get to know what was happening in such small circles by just talking to other members of the circle. Everybody knew selected bits about other members of these groups and related ones, either directly, or through other fellow group members. Everybody knew, more or less, whom to trust and whom or what to question. The flow of information was implicitly filtered and annotated with meta-information such as trustworthiness, reputation, or influence. The power of information transfer was then in the hands of all (or most) people.

As societies grew larger, the appearance and then growth of media took over the role of primary information channel. Media, first appearing in printed form, then in broadcast and electronic formats, gained tremendous power by being the 'gatekeeper' of information, with many incidents of 'tweaking' information for the benefit of higher sources of power. It then became apparent that the power of controlling information flows is immense. In recent years, however, we have been witnessing the potential of power being transferred back to the people who are getting more and more opportunities to access information directly from primary sources, with geographical distance being no obstacle. We label these new forms of communication and interaction 'e-word of mouth'. Often, it is also referred to as 'Social Media', respectively communication via Social Networks.

Social media incorporate web technologies that mediate human communication. They take on many different forms, such as, blogs, wikis, videos and photo sharing platforms, or social bookmarking sites. Functionalities on such platforms include content file uploading, publishing, instant messaging, commenting, and recommending, to name but a few. Highly successful social networking platforms like Facebook and Twitter have more than one of the above functionalities attracting audiences which, in terms of member numbers, compare to the largest countries in the world. (If Facebook were a country, it would be the third largest in the world, after China and India, with its roughly 1 billion members.)

Nowadays, almost everything and everyone can be found in or via Social Media. It takes a lot of effort though, and often it is even impossible, to distinguish the useful information from the noise; it is like trying to find the tiny needles in very huge haystacks. This need has been acknowledged by relevant research communities and several initiatives that have lately appeared. Many of these focus on the automatic discovery of information by adopting semantic search and retrieval technologies, adapting this to the particularities of Social Media content. What has not yet been achieved, however, is to discover and utilise the hidden social values that were implicitly exploited in former small 'physical' societies.

In Social Media we do not only have bare content; we also have to deal with interconnected sources. Furthermore, we have to handle and interpret interactions between sources. And there is much more information available that indicates the context within which the content is being used and the context in which these interactions are taking place. Complete models of small and large societies, interacting vividly in real time, are engraved and well hidden in the content that is residing in Social Networks. If we can manage to decipher these interactions we will be able to reveal much more than the direct meaning, or what is evident "on the surface". Further to discovering what is being said about an unravelling event, for example, we will be able to tell how trustworthy a piece of information is. In addition to discovering what people are saying about a product or service, we will be able to predict how influential these people are and how much this can affect the reputation and marketability of the product or service. We are entering a stage that allows us to automatically judge the quality of content, and even predict future trends.

Along these lines, the REVEAL project aims to advance the necessary technologies for making a higher level analysis of social media possible, thus enabling users to reveal hidden 'modalities' such as originality, trustworthiness, reputation, proximity, influence or credibility of information etc. In order

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 5 of 43

to achieve this, the work and efforts of the consortium during the current reporting period focused along the following axis:

- User requirements and regulatory framework for the REVEAL platform

- Realisation of the REVEAL platform, consisting of:

  o Contributor modelling

  o Multimedia analysis and indexing

  o Context-centric information analysis

  o Modalities analysis framework

  o Framework integration and REVEAL applications

- Evaluation and impact analysis of REVEAL solution

- Training, dissemination and exploitation activities

, which are described in detail in the following sub-sections.

1.1 Work performed and results achieved during the first year of the project

User requirements and regulatory framework for REVEAL platform

- literature study on the end-users' needs related to the verification process of content in news in general, and dealing with information residing in social networks in particular, and understanding of the current challenges in verification processes;

- conduction of interviews for qualitative and quantitative analysis of user requirements;

- in-depth comprehension of how end-users utilize social media for (i) verification purposes, (ii) how the verification process looks like, and (iii) which platforms and processes are used for verifying content;

- conduction of a market analysis of already existing platforms and the various features;

- compilation of mini-scenarios based on the findings through the literature study, the interviews, and the market analysis;

- alignment of the defined user requirements with the s/w modules to be developed;

- analysis and description of the applicable legal framework in the area of privacy and data protection;

- monitoring of legal and policy developments in the areas relevant for the REVEAL project (data protection, media law, intermediary liability) with a view to providing critical commentary in scientific and policy-oriented fora;

- provision of assistance to the partners in compliance with their formal obligations related to the data processing activities in the project;

- specification of the legal requirements for processing of personal data, conduction of a legal analysis of the technical modules to prevent any infringing activities within the project, and articulation of a practical legal recommendation guide for the mini-scenarios;

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 6 of 43

- establishment and first contact with the legal advisory board committee members;

Realisation of the REVEAL platform

- **Contributor modelling**

  o definition and partial implementation of a software architecture for the contributor modelling;

  o identification and collection of suitable data sets for initial experiments and evaluation purposes;

  o conduction of a literature review on gamification and crowdsouring, and preliminary prototype implementations of gamification modules deployed in SAG communities;

  o identification of contributor features and development of individual approaches for detection of contributor related modalities and features;

  o implementation of early prototype functionality for topic based influence, user role and thematic user classification, multipartite communities, as well as a module for topic detection;

  o definition of the contributor modelling s/w modules, their envisaged functionalities and dependencies on the other REVEAL modules, and the respective modules' input/output formats.

- **Multimedia analysis and indexing**

  o definition of the multimedia indexing framework requirements, and design of a first version of the multimedia analysis indexing architecture, its components and methods;

  o evaluation of open-source web crawlers on the task of gathering large amounts of images from news sites or other sources, as well as extension and customization of the chosen solution (BUbiNG) according to the elicited requirements;

  o development of a supervised method for performing local graphic and font overlay detection in order to improve near-duplicate image search, increasing image reposting detection accuracy;

  o comparative evaluation of the existing image similarity search method (SURF+VLAD and PQ) with the proposed extension on standard image datasets;

  o preliminary development and experimental evaluation of the concept detection, geographic concept detection and image clustering modules based on state-of-the-art techniques on standard datasets;

  o implementation of a preliminary prototype incorporating multimedia item text-based, visual indexing and clustering, the results of which were exposed through REST services. A web UI was built to demonstrate the implemented functionality. The test dataset used for the demonstration was the SNOW 2014 Data Challenge dataset;

  o review of a number of state-of-the-art HTML segmentation approaches, and comparative experimentation and selection of the HTML segmentation approach, for the cleaning of the crawled web pages (extraction of main text and images);

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 7 of 43

- o setup of the linguistic pipeline for the text analysis (tokenization, sentence splitting, part-of-speech tagging) and experimentation and first development of the Named Entity Recognition module based on the Stanford NLP library;

- o review of state-of-the-art relation extraction approaches and experimentation with a first approach on a subset of the SNOW 2014 Data Challenge dataset;

- o review of existing image forensic techniques and definition of a list of manipulations, which would be valuable for the journalistic REVEAL use cases;

- o processing of a recent version of the Common Crawl dataset, using the AWS Elastic Map Reduce (EMR) infrastructure (50 nodes), for the extraction of web-scale image statistics and the identification of an extensive collection of news image URLs;

- o definition and implementation of a unified model for socially interconnected multimedia-enriched objects.

- **Context-centric information analysis**

  - o definition of an architecture for context extraction and analysis;

  - o identification and collection of suitable data sets for initial experiments and evaluation purposes;

  - o review of related work and state-of-the-art, development of individual approaches for geospatial, social and topical content extraction methods, and implementation of early prototype functionality for selected methods;

  - o generation of first theoretical and practical results of the individual approaches;

  - o investigations of public knowledge from the Linked Open Data cloud for the purpose of a semantic context analysis;

  - o definition of the context-centric information analysis s/w modules, their envisaged functionalities and dependencies with the other REVEAL modules, and the respective modules' input/output formats.

- **Modalities analysis framework**

  - o investigations of project datasets with respect to trust and credibility modalities leading to definition of JSON annotations;

  - o technology evaluations related to the modalities analysis framework feeding into architecture choices and screen mockups of possible visualizations for end user feedback;

  - o initial definition of the modalities analysis framework architecture and JSON annotation metadata templates;

  - o preliminary software development work on a knowledge-based trust and credibility model, and situation assessment framework components;

- **Framework integration and REVEAL applications**

  - o definition of the architectural integration requirements of the REVEAL framework and provision of the first draft of the architecture design, consisting of:

- classification of the components and the relations between them and the intended technologies (web app);

- finalisation of the REVEAL technology stack;

- diagram representation of the components and their technologies.

  o definition of the approach to be followed for the software development and integration, involving:

    - research for options about NoSQL repos (mongodb, couchDB, Cassandra, Hbase, terracota);

    - research and local testing of Storm + RabbitMQ examples;

    - research about continuous integration solution: Hudson, Maven, Sonar;

    - research and testing of Docker + Vagrant for remote setup of storm clusters;

    - research and evaluation of Lambda architecture and possible adaptation in REVEAL;

    - preliminary study on synchronization and orchestration of non-storm modules, and preparation of the integration plan and quality assurance procedures to be followed for s/w development and integration.

Evaluation and impact analysis of REVEAL solution

- preliminary developments of the evaluation plan;
- investigation of new approaches for aggregating data from social media and journalists.

Training, dissemination and exploitation activities

- establishment of the dissemination strategy, identification of relevant stakeholders to the project, and documentation of the dissemination actions (along with a set of evaluation metrics) that need to be conducted for raising project awareness;

- design and creation of promotional material such as the project's graphic identity, REVEAL logo and avatar, brochure, and a poster;

- establishment and maintenance of project portal and social media accounts (Twitter, youTube, Linkedin) with information pertinent to project advances and news relevant to the project's research fields;

- publication of 17 scientific research papers and several press releases in different countries for raising the project awareness;

- provision of a set of training activities under the umbrella of the well recognized International Research Centred Summer School (IRSS2014);

- attendance of events relevant to the project, presentation of intermediate project results, and distribution of promotional material;

- preliminary identification of REVEAL exploitable tangible assets;

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

- conduction of market watch on the latest trends in social media, a market analysis in the news media and enterprise fields, and a SWOT analysis for the integrated REVEAL platform;

- identification of candidate stakeholders, and relevant and competitive products available on the market.

- documentation of preliminary individual exploitation intentions.

1.2 Reference to the project public website and project logo

Project website URL: http://reveal-project.eu

1.3 Next steps for second year of the project

User requirements and regulatory framework for REVEAL platform

- monitoring of market developments and literature, focusing on new features provided by existing verification platforms as well as emergence of new platforms, services and initiatives;

- alignment of REVEAL modalities with the envisaged modules to be developed so as to ensure that the use cases are in accordance with the defined user requirements;

- continuous assessment of end users' needs by further interviewing end-users;

- continuous legal analysis of the technical modules to prevent any infringing activities within the project, along with legal checks especially with regards to privacy issues;

- provision of assistance to the partners in compliance with their formal obligations related to the data processing activities in the project;

- monitoring of legal and policy developments in the areas relevant for the REVEAL project (data protection, media law, intermediary liability) with a view to providing critical commentary in scientific and policy-oriented fora;

- descriptive and critical analysis of the applicable legal framework in the area of intermediary liability.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 10 of 43

Realisation of REVEAL platform

- **Contributor modelling**

  o development of methodology and respective s/w module for supporting trust elicitation between users, community based user profiling, and estimation of presence modality;

  o structure-based user profiling development and refinement;

  o community characterization and community evolution;

  o Integration of contributor modelling s/w modules with PServer and graph databases.

- **Multimedia analysis and indexing**

  o extension and refactoring of existing components (web crawler, HTML segmentation, NER, concept detection), and implementation of additional components of the base multimedia analysis and indexing toolbox (geotagging);

  o further extension of the image verification corpus: https://github.com/MKLab-ITI/image-verification-corpus and consideration of means/tools to make its extension and maintenance more streamlined;

  o integration of the real-time analysis modules to the modalities analysis framework using Apache Storm and RabbitMQ;

  o further analysis and exploration of the Common Crawl dataset, with a focus on images and image metadata;

  o further experiments on image clustering and relation discovery components. Image clustering will be combined with focused crawling processes to create image clusters on demand using open web content;

  o selection and implementation of first image forensic analysis methods and execution of preliminary experiments;

  o Review of the state-of-the-art for text stylometry and experimentation with a first approach.

- **Context-centric information analysis**

  o extension, refinement and implementation of individual approaches for geospatial, social and topical context extraction and analysis;

  o integration of geospatial context extraction and analysis methods into a first demo prototype using the stream-oriented processing architecture;

  o subsequent integration of modules for social and topical analysis;

  o development of first approach for joint semantic analysis of contexts.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 11 of 43

- **Modalities analysis framework**

    o dataset acquisition for the purposes of evaluation work. This includes working with end user partners to acquire ground truth data for specific scenarios and the technical community to get benchmark labeled data;

    o scientific evaluation of the trust and credibility models developed leading to publication of results in peer reviewed scientific papers;

    o development of framework software including models, situation assessment framework and visualizations

- **Framework integration and REVEAL applications**

    o finalisation of the REVEAL architecture;

    o delivery of first prototype implementations of the REVEAL integrated platform, News and Enterprise applications.

Evaluation and impact analysis of REVEAL solution

- assessment of the impact of new technologies on existing rules in the areas of privacy and data protection, and intermediary liability;

- formulation of a first draft of policy recommendations to address the identified obstacles in the areas of privacy and data protection, and intermediary liability;

- development of the user evaluation plan;

- provision of a first set of new methods for capturing user behavior and experience.

Training, dissemination and exploitation activities

- continuous reporting on the website about project achievements, market developments and other aspects of relevance for REVEAL work;

- increase the project awareness via the publication of scientific papers, press releases, and attendance of relevant to the project conferences and events relevant to the project;

- establishment of a vivid community network for REVEAL including experts, industry and researchers worldwide;

- finalisation of REVEAL exploitable assets, and preliminary identification of licensing scheme and associated IPRs.

- update of market analysis and enhancement of business exploitation plan.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 12 of 43

2   Project objectives for the period

The most important project objectives for the 1st year of the Reveal project have been reached and are the following:

User requirements and regulatory framework for REVEAL platform

- conduction of a user requirements elicitation process and designing and categorization of mini-scenarios based on the literature study, different technical modalities, market analysis and interviews;

- gaining an overview of the market and reviewing the most popular products and services as well as features offered to end-users;

- analysis and description of the applicable legal framework in the area of privacy and data protection, along with the specification of the legal requirements for processing of personal data;

- legal analysis of the technical modules to prevent any infringing activities within the project;

- provision of assistance to the partners in compliance with their formal obligations related to the data processing activities in the project;

- establishment of the legal advisory board committee.

Realisation of REVEAL platform

- **Contributor modelling**

    o identification of data sources and methods for obtaining social network data from twitter, stackexchange, and youtube. Typically the data contain structural (e.g. who follows whom) and content information (e.g. posts);

    o definition of the initial form of contributor profiles and partial implementation of a contributor modelling architecture;

    o conduction of a literature review on gamification for enhancing the contributor participation in REVEAL platform;

    o preliminary study on the analysis of communities around people and content;

    o development of initial approaches and preliminary prototype implementations for the estimation of contributor modalities.

- **Multimedia analysis and indexing**

    o definition of the multimedia indexing and search requirements, the corresponding architecture and the involved components, along with a list of specific methods and services exposed by the latter;

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 13 of 43

- o implementation and evaluation of a number of techniques for multimedia indexing and search approaches, in particular for web image crawling, HTML segmentation, visual feature extraction, near-duplicate search, concept detection, Named Entity Recognition, image clustering and relation discovery;

- o implementation of a preliminary prototype (featuring a set of REST services and a web UI) demonstrating the implemented services on a static collection;

- o extensive study and categorization of image forensic analysis methods.

- **Context-centric information analysis**

  - o definition of the context-centric information analysis architecture for extraction and analysis of context in social media which interfaces with the broader analytical scope of the modalities analysis framework;

  - o development of initial approaches for geospatial, social and topical content extraction and analysis, along with the provision of first theoretical and practical results;

  - o draft of methods for semantic context analysis, including an analysis of public sources for extending context information.

- **Modalities analysis framework**

  - o conduction of technology evaluations for trust and credibility modeling, situation assessment framework, and visualization for the modalities analysis framework;

  - o definition of the modalities analysis framework architecture

  - o provision on an analytic report on metadata template definitions

- **Framework integration and REVEAL applications**

  - o delivery of the first version of the REVEAL architecture and specifications;

  - o definition of the integration plan and conduction of preliminary activities necessary for the establishment of the integration environment in terms of s/w and h/w resources.

Evaluation and impact analysis of REVEAL solution

- preparation activities for the establishment of the evaluation and impact analysis framework.

Training, dissemination and exploitation activities

- establishment of the dissemination strategy;

- identification of target audiences and definition of a preliminary market analysis and exploitation business plan.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 14 of 43

- development and maintenance of the official project's portal and social media accounts;

- design and delivery of promotional material, publication of scientific papers and press releases, and attendance of important dissemination events for raising the project awareness.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 15 of 43

3 Work progress and achievements during the period

3.1 WP1: User Requirements and Regulatory Framework

Work-package objectives

The overall objectives of WP1 concern the establishment of the foundations for all subsequent development work. This involves a recap of the assumptions that were made in the proposal submission process and period, and a revisit of the market and stakeholder analysis that has been conducted during the proposal stage, so that these are updated for reflecting the recent trends and developments with regard to existing tools, services and market needs. In the context of this work-package two detailed requirements analysis will be conducted, one in the area of journalism, and the other in the field of enterprise networks. In addition, the relevant regulatory framework will also be examined to ensure that the development will be in line with current regulations and legal practices on personal data protection. Based on the outcome of work conducted in WP1, developer partners will propose their final system specifications and architecture design.

Progress towards objectives

Task 1.1: Requirements analysis in the journalism field

This task started at project month M1 along with the project initiation and focused on the requirements elicitation process in the journalism field which involved (i) a set of expert interviews with journalists, editors and people working in media organisations, (ii) a literature study to outline facts about already existing fact-checking practices used in newsrooms and by journalists, and (iii) a market analysis in order to identify existing verification platforms and services. All of the above mentioned activities were performed in a collaborative environment among DW (leader of this task and journalist pilot of REVEAL project), SINTEF, and ATC.

The outcome of the interviews, market analysis, and the literature study was summarized in mini scenarios which focus on very specific end user challenges that can be seen as typical journalistic workflows. These mini scenarios were further analysed in specific requirements, so-called user stories and categorised based on what modalities they are planning to reveal. These requirements were prioritised using a ranking scheme according to their importance and mapped with the work packages and tasks responsible for their implementation. Results of the activities conducted under this task were reflected in Deliverable D1.1.

Task 1.2: Requirements analysis in the enterprise field

Similarly to the previous task, this one started at project month M1 and ended at M6 with the objective of targeting the analysis and definition of the user requirement in the enterprise field. In order to meet this objective SAG (leader of this task and enterprise pilot of REVEAL project) concentrated on a different approach than DW did in task T1.1, capitalizing upon a set of user focused-groups. This involved discussions and brainstorming sessions on studying observations of user behaviour in a team of experts from the ARIS Community[1] that led to the preliminary identification of a number of research questions and the ideal stakeholder groups that should be targeted in the context of this exercise. As a second step, efforts were intensified by proceeding to

---

[1] ARIS community is part of TECHCommunity maintained by SAG featuring approximately 322.050 of active users.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 16 of 43

the conduction of numerous interviews with members from the stakeholder groups that were identified.

Following the input that was collected throughout the surveys that were conducted in the context of this task and taking into account the nature of a software company, a set of mini scenarios that are tightly focused on specific end user challenges in the enterprise community area were defined. Based on these scenarios, a list of detailed user requirements were discussed, consolidated, and categorised in collaboration approach among SAG, SINTEF, and ATC. As done in the journalist case, those requirements were then mapped to the architectural design, and the work packages and tasks responsible for their implementation. Results of the activities conducted under this task were reflected in Deliverable D1.1.

Task 1.3: Legal/regulatory requirements identification & analysis

While the first two tasks of Work-package one focused on the requirements elicitation process of journalism and enterprise field, this task concentrated on the establishment of the regulatory framework to safeguard the implementation and exploitation activities within REVEAL in terms of legal perspective. Towards these lines, during the current reporting period the relevant legislative framework in force within EU to data privacy and ethics was analysed and studied. Along with the mini-scenarios that were derived out of tasks T1.1 and T1.2, and the a close collaboration of KU Leuven (legal expert and leader of this task) with the pilot and technical partners of the consortium a set of practical legal recommendations to guide the development activities of the project and ensure the REVEAL platform's legal-compliance were developed.

Apart from the regulatory framework that will continue to evolve until the end of the project duration, preliminary results of which were documented in Deliverable D1.2, other activities that were performed under this task concerned (i) the assistance of the technical team members on their formal obligations related to the data processing within REVEAL, (ii) the preparation and authoring of the necessary documentations for legal compliance purposes in the context of this project e.g. controller and processor agreements, Non-Disclosure Agreements, as well as (iii) consultancy support to the technical teams that needed to contact their national Data Protection Authorities.

Task 1.4: Consolidation of results and knowledge transfer

This task started two months earlier (at project month M1) than it was initially planned. The activities conducted within the first year of the project concentrated on the following three main pillars:

- *Alignment*: for ensuing that the project use cases (journalism and enterprise) and requirement analysis are inline;

- *Knowledge transfer*: for assuring a proper knowledge transfer and establishing a common and coherent view of project objectives within the technical teams;

- *Maintenance*: for sustaining and updating the user requirements and their alignment to the legal framework throughout the project lifetime as the project matures.

This task acted as the 'bridge' between the user requirements tasks T1.1 and T1.2, and the legal one T1.3 that brought closer the legal experts, pilot users and technical people of the consortium with the objective of ensuring that the user requirements that are defined in the former tasks are aligned with the legislative framework in force and also are possible to be realised in the context of the project.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 17 of 43

The thorough review of the defined user requirements and legal recommendations did not only contribute to the development and finalization of the Deliverables D1.1 and D1.2 but also provided input to the architectural design and risk management in terms of technical perspective. This was due to the analysis that was performed on the initial list of user requirements, which led to a preliminary identification of modules (that was served as input to WP6) that could support the defined mini-scenarios (which will be evaluated in the context of WP7). Possible overlaps between components that would realize the modalities of the REVEAL platform, missing functionalities and foreseen risks stemming from the research involved in the context of developing the latter were also part of this task.

Deliverables and milestone list for WP1

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D1.1	Requirements analysis and specifications	1	ATC	R	PU	M6 = Apr. 2014	Y	5 May 2014
D1.2	Legal/regulatory requirements analysis	1	KU Leuven	R	PU	M6 = Apr. 2014	Y	5 May 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
1	Project setup	1, 8, 9	INTRA	M6	Y	April 2014
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

3.2 WP2: Contributor Modelling

Work-package objectives

The overall objectives of WP2 concern the discovery of contributor models across different social networks taking into account the contributor activity, and the structure/characteristics of the social networks in which the contribution is effectuated. The contributors' profiles across different networks will be unified and contributors' models will be derived. Moreover, in the context of community and influential user detection an analysis of the social network(s) will be conducted. Finally, the trust level of each contributor will be derived and used in modelling the spread of information.

Progress towards objectives

Task 2.1: Contributor profiles

This task started one month earlier (at project month M1) than it was initially planned by collecting a number of social network data sets for the experimental purposes of Work-package two; and the rest of the technical WPs. A number of social media web crawlers were also employed for enriching the existing data sets incrementally, and a study on ontologies pertinent to the social media domain was also made. As a result the SIOC ontology was chosen as an ideal case for providing access to linked data, forming in this instance the 'bridge' among the data that need to be exchanged among the modules to be produced out of Work-packages WP2 and WP4.

The initial form of contributor profiles for the journalism and the enterprise domain was also defined, and preliminary prototype profiles were developed and deployed on 'PServer' which substantially facilitates the user modelling. Moreover, several steps on multi-label user classification were undertaken with the aim of revealing the user interests. In specific, a method was developed that makes use of the underlying network of online interactions (e.g. mentions, retweets) between users and a small set of "seed" users (i.e. users whose labels/classes are known) to predict the labels of the unknown users. It is worth mentioning that the particular method was evaluated on a number of datasets and was found to outperform several state-of-the-art methods.

In addition, effort was placed on investigating different gamifications approaches which may suit and be employed in the REVEAL enterprise application with the objective of strengthening the user engagement and therefore contribute to the enhancement of modalities' detection. Finally, an initial architecture of the modules that would constitute the integrated REVEAL platform under the framework of WP2, and the interfaces for enabling the interlinking of those components with other ones developed by the rest of the technical Work-packages (WP3 to WP5) was defined capitalising upon the PServer infrastructure. This was also served as preliminary input to task T6.1 of WP6 for delivering the architectural design of the integrated REVEAL platform.

The outcome of the activities conducted under this task was documented in Deliverable D2.1 that was submitted at M10.

Task 2.2: Analysis of communities around people and content

This task started at the second semester of the current reporting period focusing on the high level definition of the various modalities that should be supported by WP2 for characterising a contributor (such as popularity, influence, presence, trust, history), and on the identification of the contributor alignment to the topic of discussion within a community. Throughout the active period of this task, a number of modules where designed and developed to either estimate directly the defined contributor's modalities or to support indirectly their detection by capitalizing upon content and/or structural related features of the underlying social network.

In specific with respect to the content feature, research was made on developing a novel language-agnostic topic detection method based on suffix trees, capable of detecting the topic of a contributor's conversation and diversions from that topic that could contribute to the discovery of provenance information. From the structural relations perspective, other modules where also designed and implemented for the discovery of contributors' roles (e.g elitist, taciturn etc.)

In addition, an influence detection module based on 'supervised random walks' techniques was produced for estimating the influence of contributors according to a-priori information, their properties, content produced and interactions of the former. Finally, during the lasts months of the current reporting period emphasis was given on experimenting and preliminary implementing a multipartite community detection module based on a density technique that will enable the detection

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 19 of 43

of communities in social networks which comprise the following entities (or partities): users, tags and resources (e.g. URLs).

As in the case of task T2.1, the outcome of the activities conducted under this task were documented in Deliverable D2.1 that was submitted at M10.

Task 2.3: Contributor behaviour analysis

Task 2.3 started in the last quarter of the current reporting period at project month M10. Initially, a review of the relevant literature on trust was conducted. This review comprised a collection of different notions of trust in social networks and other online platforms as well as a study on features which are considered relevant to trust. In a second step, potential datasets for trust analysis were investigated and feature engineering was commenced on selected suitable datasets.

Moreover, some preliminary work was undertaken in the specification of the contributor's reputation modality, as well as in the definition of the alignment module that will determine the match of a user, given his/her posts, with a query. Work was also performed for defining the source finder module, which will enable the discovery of the original posts that sparked the creation and evolution of a highly discussed topic.

Finally, a review on methods for duplicate account detection across multiple social networks was undertaken and an initial list of a number of social networks were identified which will be used for experimental purpose; a preliminary study on obtaining relevant data was also undertaken.

Deliverables and milestone list for WP2

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D2.1	Contributor modelling requirements and baseline approaches	2	NCSR'D'	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

3.3 WP3: Multimedia Analysis and Indexing

Work-package objectives

The overall objectives of WP3 concern the extraction of hidden modalities from the content of media items shared on social media platforms. More specifically, WP3 aspires to develop a toolbox for facilitating the indexing of user-generated content with the dual goal of (a) discovering high-quality and trustworthy media content, and (b) detecting fraudulent, misleading and duplicate content. Thus, research and development in this work-package targets the following outcomes:

- a base multimedia indexing toolbox enabling large-scale content-based search and filtering capabilities on large amounts of multimedia content collected from OSNs;

- novel scalable approaches for linking online content with the goal of analysing content creation, modification and spreading across OSNs;

- hidden modality extraction approaches, in particular approaches dealing with text stylometry and with the detection and analysis of digital image manipulations.

The planned indexing approaches will be based on text, visual and hybrid (making use of both text and visual content) analysis methods. Due to the complexity of the problem at hand, the output of the envisioned methodologies is expected to be combined with user relevance feedback acquired as part of the end user pilot applications (WP6) so that human judgement is included in the content analysis loop to ensure verification of the produced results.

Progress towards objectives

Task 3.1: Multimedia indexing toolbox

During the first year of the project, task T3.1 was the longest running and most intensive task of WP3, starting from project month M2. On a collaborative approach with the geo-spatial context extraction task of T4.1, the modalities analysis framework of WP5, and the framework integration and REVEAL application of WP6 a fist set of fundamental architectural decisions were taken in relation to the planning and guiding of the respective research and implementation activities that need to be conducted within those tasks/Work-packages. Particularly, for the set of real-time analysis modules of WP3, it was agreed to use a real-time, distributed stream processing approach based on Apache Storm[2]. Exceptions to this style of processing include components that by their nature require batch processing such as the image clustering, the relation extraction and stylometry modules. In addition, it was a common decision to use RESTful web interfaces for the predominant multimedia indexing functionalities (related to the exposure of the extracted results and search services), which could be leveraged directly by the end users and/or other components of the REVEAL platform.

With respect to the visual content analysis conducted under the framework of this task, for the purpose of uncovering the provenance of multimedia content (e.g. to detect possible re-postings of the same image on the Web) several third-party reverse image search tools were evaluated and an existing near-duplicate image search implementation based on SURF descriptors was extended, tested and evaluated on several datasets. The implemented extensions aimed at improving the algorithm performance in the presence of overlaying fonts or banners on a given image. Furthermore, extensive concept detections experiments were made using a semi-supervised approach based on Approximate Laplacian Eigenmaps in order to achieve large-scale multimedia

---

[2] http://storm.incubator.apache.org

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 21 of 43

annotations. This investigation was performed in collaboration with WP4, as the output of the concept detection module will be of considerable value for the geospatial context extraction in WP4 (for geographic concepts). In terms of geotagging multimedia content, text-based techniques as well as techniques based on the visual content were also studied.

In the text content analysis field, an initial attempt to HTML segmentation with the objective of extracting the relevant text and the title of an article, as well as related images from crawled web pages was made. This involved the investigation and evaluation of the relevant state-of-the-art approaches, and the experimentation and setting up of the linguistic pre-processing pipeline that will be used for the needs of WP3 modules. The purpose of the employed NLP pipeline processing mechanism is to process texts (e.g. tweets and posts) which are collected from the social media crawler in order to gain a better understanding of the linguistic and structural information associated with the collected data, as well as to enable the extraction of semantic information from that data that may enhance the precision of modalities' detection. Towards these lines, a prototype implementation of the Named Entity Recognition module was implemented, capable of identifying people, places and organization names as well as dates within texts by exploiting pre-processing techniques related to tokenization, sentence splitting, part-of-speech (POS) tagging, shallow parsing, lemmatization or stemming, and abbreviation detection.

For the needs of this task, a large-scale image crawl on a specific set of web-sites and the Web overall was a necessity. For that purpose, an extensive comparison of several available open-source crawlers was done, and a custom image crawler based on the BUbiNG[3] web crawler was implemented. In addition, in order to draw insights about the role and function of images on the Web, the Common Crawl dataset was processed and analysed. Finally, it should be noted that although the project is still at an early stage of s/w development, WP3 managed to produce several tangible assets. A working prototype of the multimedia indexing web-service is already in place and most of the WP3 modules are at a mature state of development ready for integration under the context of WP5 and WP6. In terms of specification, an extensive architectural description of WP3 framework, as well as output examples in the form of JSON mark-up may be found within Deliverable D3.1 (and parts of them in D4.1 and D5.1) that were submitted at M10.

Task 3.2: Multimedia relation discovery and linking

This task began at project month M6 and focused on the relation extraction and clustering of multi-media content.

In the context of clustering, preliminary activities concentrated on the investigation and evaluation of existing clustering techniques and their applicability to REVEAL needs for (i) extracting highly similar groups of images from large streams of social media content in real-time in order to detect emerging news stories - something that is of great importance to the journalism domain, and (ii) utilising highly sophisticated clustering methods that aspire to disclose information about the provenance of multimedia content. In terms of implementation, a simple incremental clustering technique was developed and integrated in the first working WP3 prototype, whereas an initial definition of the output format of the clustering module was concluded and specific directions for future work were identified.

In a similar way, extensive investigation of existing approaches for relation extraction took place. Several approaches were evaluated on social media data, and specific emphasis was given on the informal and unstructured nature of social media language; something which has not yet been brought to the attention of other research communities working in this field. A number of experimental tests were conducted with the objective of extending existing approaches to the nature

---

[3] Description on: http://dl.acm.org/citation.cfm?id=2577304, code on: http://law.di.unimi.it/software

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 22 of 43

of social media text for increasing the accuracy of the employed algorithms, and a first prototype implementation was derived.

Task 3.3: Multimedia forensic analysis

Although this task started at the last quarter of the current reporting period substantial progress was made. The activities conducted under this task involved an extensive investigation of the state-of-the-art techniques in forensic analysis; something which is still in progress.

Image forensic analysis is a broad research area which takes advantage of several features of the underlying image structure in order to draw conclusions concerning the authenticity of an image in question. Up to now a number of different techniques were studied ranging from pixel-based approaches that detect anomalies at the pixel level to format-based ones that focus on the statistical correlations introduced by a lossy compression format like JPEG. More advanced once were also investigated such as camera-based techniques for exploiting irregularities of the camera lens. Geometric irregularities, like the abnormal position of a certain object relative to the camera or an anomalous interaction among light and the objects in a digital image, which can also be used to infer whether or not a particular picture was maliciously manipulated.

Throughout this study, it became clear that there is no unified framework or approach suitable for detecting a multitude of manipulations in near real-time. Furthermore, most of the existing techniques that have been researched are not tested against large and diverse image datasets (and definitely not from images shared on social networks). Therefore, a secondary goal of the performed investigations was to determine the extent at which WP3 could produce an extended test dataset and evaluation process for image manipulation detection methods. In collaboration with DW a list of manipulations were identified, whose detection would be valuable from a journalistic perspective, and as a short-term goal of task T3.3 it was decided to target at delivering a high accuracy in the detection of some of these manipulations and possibly introduce more challenging ones in the future.

Finally, in the light of stylometry a first investigation of the respective state-of-the-art approaches was initiated. Although most of the current techniques are suitable for large texts such as books, limited focus is given to social media content due to the associated challenges related to the unstructured language used and the brevity of texts. In this instance, the aim of this task for the coming months is to work on and develop new methods for discovering the important features of the different writing styles in social media.

Deliverables and milestone list for WP3

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D3.1	Multimedia indexing requirements and baseline approaches	3	CERTH	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 23 of 43

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

3.4 WP4: Context-centric Information Analysis

Work-package objectives

The overall objectives of WP4 concern the provision of a toolkit for analyzing the different dimensions of context of both contributors and content items, and the coherence relations between context items with reference to the social, topical, and geo-spatial context dimensions. The meaning, influence, significance, and authoritativeness of a particular content item will also be considered. Apart from all these dimensions, and the methods of extracting those from social network content items, this work-package will also investigate general semantic analysis methods that are to be applied for fusing individual context information into a general picture.

Progress towards objectives

Task 4.1: Geo-spatial context extraction

Work in task T4.1 commenced at project month M2. Due to the fact that this was the first task to start in WP4 within the current reporting period, all of the involved technical partners have used the context of T4.1 to conclude on decisions related to software architecture and design aspects of WP4 modules.

Similarly to WP3 and in close collaboration with ITINNO it was agreed to follow a stream oriented processing paradigm utilising the Apache Storm framework and a RabbitMQ[4] message bus for ensuring the scalability aspect of WP4 analytical modules and a coherent integration with the modalities analysis fusion framework of WP5. The first set of material that was produced out of this task and shared among the technical team members was a documentation of the WP4 platform technologies, a concise technical manual featuring architectural details, and a set of example code samples. For the purpose of running analytics and drafts of algorithms early on, a set of data sets suitable for geospatial context extraction was investigated and compiled. The identified data sets concerned different types of media, formats, and scenarios for geospatial context extraction and were shared with the rest of the partners. Most of the data sets have already been established in a scientific context and were, thus suitable for comparison to alternative analytical approaches.

In specific, the research work on the methods for geospatial context extraction was pursued in several directions. Following different types of media, data sources and the availability of voluntarily provided spatial information, the methods covered a wide range of approaches. As it was concluded,

---

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

named entity matching over information extracted from OpenStreetMap[5] provided fast and high-precision covering of locations in text based contents, but limited to a pre-defined and restricted set of regions. On the other hand, region-aware topic models were found to be more coarse grained, but restrictive to geospatial information which may be inferred from pure text contents. As for geospatial context information that is possible to be extracted by multimedia analysis techniques emphasis will be given on visual features and visual similarity in images and photos. In addition, on an attempt to extend and enlarge a given geospatial context a further study was made on the exploration of publicly available geospatial information maintained in the Linked Data cloud.

As a result of the activities conducted under this task a set of preliminary prototype modules were produced with clearly defined input and output data formats based on JSON; dependencies and interaction between those modules were also expressed in a diagrammatic and document form. The output of this task and the subsequent one (T4.2) were documented in Deliverable D4.1 that was submitted at M10.

Task 4.2: Social context extraction

This task started in the middle of the current reporting period at project month M6 by performing a literature review of the relevant state-of-the-art which led to (i) an overview of features for social context which have been investigated previously and are established in the field, and (ii) a collection of different resources of publicly available data sets which may be used for initial experimentation purposes.

The developed list of features comprised a wide range of interactions to cover social context among users, between a user and content items and interaction with the focus set on content items. Features of social context among users cover mainly observations of interaction between users. These interactions may be direct (e.g. friendship relations, follow relations) or indirect via additional entities (e.g. responses to content items, co-authorship, mentioning), whereas features of social context between users and content items build on structures such as reply-graphs, communities around content items and information cascades. Although, a set of core features is generic and applicable to a wide range of online communities and social media, an extended set of features is not available in all sources or might even be community specific. Nevertheless, as these communities and their features might be of interest and relevant to the REVEAL use cases will also therefore be considered in the analysis.

In terms of implementation, initial methods for context feature extraction and analysis were developed and have already produced some preliminary results. In specific, a preliminary analysis of social context features among users was performed on several StackExchange[6] online communities, and the derived features were analysed for correlation in order to enable feature selection in future steps; this is something that is/will be used as input to WP2 modules. In addition, a first approach to track communities over time was implemented, involving an analysis of how communities evolve over time, a taxonomy of which events can be detected, as well as an algorithm for tracking and detecting the evolution events. Finally, in terms of integration efforts were also placed on the definition of the input and output format of WP4 modules. Although this is work in progress it was found to be useful as it reflected the expected functionalities that should be exposed by the WP4 modules.

---

[5] http://www.openstreetmap.org

[6] http://stackexchange.com

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 25 of 43

Task 4.3: Topical context extraction

Task 4.3 commenced at the last quarter of year one, and despite its late initiation concrete results were produced during this period.

Initially, several designs of context-aware approaches for topic modelling were drafted, and following the conceptual design and methods of the geospatial topic models (which were defined in T4.1), those models were further enhanced by introducing alternative types of context that may be included into topic models. Indicative examples include context related to contributor, community and time; nevertheless the corresponding probabilistic topic models and their design still needs to be finalized and investigated for a tractable computation of the models.

Another line of research has pursued the development of improved smoothing techniques for language models. In specific, the developed Generalized Language Model approach that was selected combines state-of-the-art Kneser-Ney Smoothing with a systematic recursive investigation of skip n-gram models, and it reflected that the obtained language models can significantly reduce perplexity over held-out text data. It particular, it was evidenced that Generalized Language Models can cope well with sparse training data which renders them suitable for application on social media with its often short text messages.

Moreover, initial steps were taken towards event detection and modelling of high- and low-level events with the working hypothesis of detecting (i) low-level events in streams of raw social media data and to define modalities as composed, and (ii) high-level events which are then detected on a stream of low-level events. Finally, it should be noted that a preliminary list of modules to be derived out of this task was compiled.

Task 4.4: Semantic context analysis

Similarly to previous task, T4.4 started at project month M10 with focus on suitable technologies relevant to semantic context, and an analysis of public knowledge available from the Linked Open Data cloud.

One line of research investigated how knowledge on the Linked Data cloud is captured and published, comprising of a number of surveys among Linked Data engineers to identify how they model knowledge, as well as analytical approaches to identify stable patterns in the data. A second direction of the performed investigations covered more practical aspects of how to access and process Linked Data on the Web. In particular, it was analysed how dynamic these public sources of knowledge are and how stable and reliable indices and data caches are in scenarios of evolving data. Results and insights obtained from these analytics and the use of semantic technologies was fed as input to task 4.1 related to the exploration of geospatial information on LOD.

Finally, as in task 4.3 an analogous level of maturity was achieved with respect to the definition of the input and output formats of the modules to be derived out of this task and the planned functionalities that may be offered.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 26 of 43

Deliverables and milestone list for WP4

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D4.1	Context extraction methods and context framework	4	UKob	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

3.5 WP5: Modalities Analysis Framework

Work-package objectives

The overall objectives of WP5 concern the utilization of the information discovered in work-packages 2 to 4 to provide estimates of 1st and 2nd level modalities, which will be combined to provide a comprehensive view of the credibility of a media source and content, and be presented to users in a unified visualization framework. In order to achieve this, WP5 will elicit framework requirements from project partners and prospective users of the system regarding the overall process flow of collecting, collating, analysing and presenting the modalities. As a result, a scalable open architecture design for presentation of results, which will be prototyped and incrementally developed over the course of the project, will be created. WP5 will be closely connected to WP1 (Requirements) initially, to the technical development work packages WP2, 3 and 4 during the course of the project, and to WP6 (the Pilot Applications work package) to ensure that the development and evolution of the architecture and its implementation takes into account the needs and feedback of the users.

Progress towards objectives

Task 5.1: REVEAL modalities modelling

Task 5.1 started in the last month of the first semester of the current reporting period and until the end of year one the following has occurred.

Initial investigations were made by technical partners with regard to the trust modelling of different datasets, capitalizing upon the SAG enterprise forum dataset, various crawled social media datasets

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 27 of 43

(e.g. Twitter, Flickr, Instagram, You Tube etc.), and publically available benchmark datasets such as MediaEval2014. Substantial focus was placed on investigating the different types of the available data and the modalities that might be possible to be extracted from them. Towards this study, the pilot partners contributed with clarifications regarding domain datasets and descriptions of existing manual approaches to trust modelling, whereas the results of these investigations were factored into the architecture and algorithm descriptions in deliverable of WP2-WP4, as well as the JSON annotation descriptions reflected in deliverable D5.1 that was submitted at project month M10.

Furthermore, throughout a set of focused-groups organised with the rest of the technical team members, ITINNO managed to create an initial set of concrete annotation examples in the form that WP5 would be expecting to receive from WP2-WP4 modules. This was something that has directly contributed to the metadata template definitions of D5.1, and the auditing of the likely modality annotations of social media content that will be available to WP5. It should be also mentioned that preliminary work on the analysis of the options for cross-checking, credibility and trustworthiness contributed significantly to the creation of the architecture descriptions in D5.1.

Finally, it is worth mentioning that work on a knowledge-based trust and credibility model with possible inference patterns from T4.1 being initially analysed in the context of the T5.1 trust modelling task has commenced and presented at the OGC TC GeoSemantics DWG meeting[7]; part of this work was also published in a technical blog article of the REVEAL portal.

Task 5.2: Metadata driven data fusion of analysis reports

This task started three months earlier (at project month M3) than it was initially planned, as ITINNO could make immediate progress refactoring previously declared background from the TRIDEC[8] project for reuse in REVEAL. In specific, the past experience from TRIDEC architecture and metadata processing has helped shape the WP5 architecture documented in deliverable D5.1, and ensured REVEAL has a scalable metadata driven data fusion framework that can meet the challenging requirements defined in WP1.

In the context of this task, a technical evaluation exercise on the Storm distributed real-time computation system and RabbitMQ message bus backbone was made in order to ensure its suitability for the real-time processing tasks of WP5, WP3 and WP4; which yield positive results. As part of this exercise, ITINNO prepared and released a storm example code package with an accompanying user manual to the rest of the technical partners dealing with real-time processing requirements (ITINNO, UKob and CERTH), which was reviewed by the latter and adopted as well in theirs WPs. It should be noted that batch-style processing requirements may also be supported by HTTP endpoints publishing to RabbitMQ.

In terms of development, efforts were also placed on the realisation of the situation assessment component which forms the backbone of the T5.2 work. In the integrated REVEAL platform, this component will be responsible for the aggregation in real-time of incremental JSON annotation reports that are published by other WP2-WP4 modules over RabbitMQ and prepare them in a form suitable for trust and credibility modelling and visualization for data analytics. Preliminary experiments were performed and currently evaluated on social media data crawled from the 2014 Ukrainian crisis period.

---

[7] http://www.opengeospatial.org/event/1406tc

[8] TRIDEC (Collaborative, Complex and Critical Decision-Support in Evolving Crises) is a project co-funded by the European Commission under the Seventh Framework Programme (#258723) that focuses on new approaches and technologies for intelligent geo-information management in complex and critical decision-making processes in Earth sciences; www.tridec-online.eu

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 28 of 43

Task 5.3: Visualisation and data analytics

This task started at project month M6 by evaluating different visualisation technologies relevant to the situation assessment that will be delivered by task T5.2. For geospatial mapping, the Geoserver and MapServer open source OpenGIS map servers were tested coupled with HTML and Javascript presentation layers (i.e. OpenLayers[9], D3[10]), whereas for social network and thematic topic visualization, experiments utilising HTML and Javascript were performed. In order to drive the evaluation of the above mentioned technologies, an example dataset was created based on the expected JSON annotations that WP5 would be expecting to receive from the other WP2-WP4 modules, as these are defined in deliverable D5.1.

As part of this work, a number of mock-ups were also produced and it is envisaged to be communicated to the rest of the partners early within the forthcoming reporting period for gathering initial user feedback before the initiation of the corresponding implementation activities on visualisation begin.

Deliverables and milestone list for WP5

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D5.1	Analytic report of metadata template definitions	5	ITINNO	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

---

[9] http://openlayers.org

[10] http://d3js.org

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

3.6 WP6: Framework Integration and REVEAL Applications

Work-package objectives

The overall objectives of WP6 concern the integration of the models of the REVEAL framework and the conduction of the conceptual modelling, design, implementation and evaluation of the two pilots in the professional sectors of journalism (for dealing with social media news content, especially focusing on verification) and enterprise community networks.

Progress towards objectives

Task 6.1: Architectural design of framework

This task started in project month M2 instead of M6 as it was initially planned. The activities conducted within the first year of the project concentrated on the designing of the REVEAL system architecture for satisfying the functional and non-functional requirements that were derived out of tasks T1.1 and T1.2 of WP1.

Following an incremental and iterative approach, task 6.1 concluded on the modules necessary to support the integrated REVEAL platform, their interconnections, and an initial set of interactions among them by means of Application Programming Interfaces. In addition, technical discussions among the technical team members led to an initial agreement on the communication protocols that need to be established and the workflows to be supported for satisfying the foreseen use cases for the whole system, along with a preliminary description of the functionalities that each of those modules should expose from a technical perspective.

The results of this task were documented in Deliverable 6.1.1 that was submitted at project month M10. Intermediate releases though were being distributed internally in frequent intervals, as the documentation was being evolving, among the consortium partners for guiding the implementation activities of the technical work-packages WP2 to WP5.

Task 6.2: REVEAL framework integration

Task 6.2 started at project month M6 by performing a literature review on the existing technologies, approaches and software tools that meet the project objectives, both in terms of preparing the working environment for the distributed technical teams involved in the development of the REVEAL integrated platform, and the realisation of the latter according to the architectural specifications set by task T6.1.

The establishment of the development environment involved the setup and configuration of a number of tools for the facilitation of code generation, continuous integration, debugging, testing and quality assurance, as well as the deployment of the initial hardware infrastructure to support the distributed and parallel execution of REVEAL procedures featuring a Storm topology. The installation and configuration of dependency s/w tools such as message broker software, database management systems, repositories, storm controllers etc. was part of the (software) infrastructure as well.

Moreover, the agile methodology for software integration that was initially described in Deliverable D9.1:'Project quality and assessment plan', along with other procedures to guide the development activities of the project, such as continuous integration and code release management policies, were defined by the leader of WP6 (ATC), discussed and agreed by all technical team members in a collaborative environment.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 30 of 43

An initial release of common structures to be used by the platform providing the first set of methods and APIs is already in place; whereas the first internal delivery of a prototype version of the integrated solution is envisaged for project month M15.

Task 6.3: News pilot application development

Task 6.4: Enterprise pilot application development

These tasks have not started yet.

Task 6.5: REVEAL framework technical validation

Task 6.5 started at the last two months of the current reporting period. During those months the plans, procedures and tools for the technical validation of the REVEAL framework were defined and set capitalising upon the guiding directions of Deliverable D9.1:'Project quality and assessment plan'.

It should be noted that the technical validation to be performed in the context of this task until the end of the project duration is not related to the evaluation of REVEAL from the user perspective (that will be conducted under WP7; involving an acceptance test), but from a technical point of view. This concerns the integration testing of REVEAL platform as a whole in order to ensure that all mechanisms and services offered by the former are functioning according to the requirements defined in D1.1: 'Requirements analysis and specifications'. Unit testing for the individual modules to be produced out of the technical work-packages is expected to be conducted at WP level by the respective module developers with the assistance of the system integrator (where needed).

Deliverables and milestone list for WP6

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D6.1.1	Architecture and specifications	6	ATC	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

3.7 WP7: Evaluation and Impact Analysis

Work-package objectives

The overall objectives of WP7 concern the understanding and identification of behaviour patterns of communities of users and their requirements, and the provision of a set of methods for capturing user behaviour and experience. In addition, WP7 regards the establishment of a test-bed and the evaluation of REVEAL (news and enterprise) applications in an iterative and user-centered manner, taking into consideration privacy and trust issues that might influence the acceptance of the REVEAL applications.

Progress towards objectives

Task 7.1: Methods for capturing user behaviour and experience

This task started one month before the end of the current reporting period (at project month M11). During this time efforts were placed on investigating new approaches for aggregating data from social media and journalists. In particular, SINTEF that leads the task focused on content analysis of websites, where journalists are discussing social media tools for research and verification, by applying Meltwater's online intelligence platform[11] which harvests billions of digital documents. Obtaining views directly from journalists and enterprise workers is expected to provide experiences in their current and emerging practices in regard to verification processes. Such information and insights may lead to interesting knowledge in how verification tools can reach a more efficient level to suit the particular needs of individual journalists and enterprise workers.

Task 7.2: REVEAL framework evaluation

Task 7.3: Legal/regulatory requirements evaluation

These tasks have not started yet.

Deliverables and milestone list for WP7

List of Deliverables

No Deliverables were due in Year 1.

List of Milestones

No Milestones were set for Year 1.

---

[11] http://www.meltwater.com

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

3.8   WP8: Training, Dissemination and Exploitation Activities

Work-package objectives

The overall objectives of WP8 concern the provision of the required tools to ensure the proper dissemination of the project results to both the scientific community and the general public, and the investigation on a number of different commercialization capabilities that may be exploited after completion of the project. WP8 will aim to reach different stakeholders and communities via a diverse number of dissemination channels, an active web portal and an online social community, and will also undertake all the necessary training activities to make REVEAL and its services easily understood and adopted by professional (as service providers) and public users.

Progress towards objectives

Task 8.1: Dissemination activities

This task started along with the commence of the project and will run until the end of the project duration. During the current reporting period, REVEAL's mission statement has been identified and the dissemination strategy has been defined, which can be divided into the phases of creating general awareness on the topic of social media verification as well as the role of REVEAL, the engagement with the verification community and having an impact in the community.

The dissemination strategy includes the identification of specific target audiences (EC – ICT community, scientific/academic community, industrial community, and the verification community at large), as well as establishing a project identity that includes the design of a logo and the definition of colour schemes for further dissemination activities, such as using these for presentations/slideshows at conferences/other events.

Furthermore, indicators for the measurement of success have been defined for the dissemination activities on different online platforms, including the website (http://revealproject.eu), Twitter (https://twitter.com/RevealEU), Slideshare (http://de.slideshare.net/revealeu) and YouTube (https://www.youtube.com/channel/UC5-mzN9eruBnIqQI0bk32mg).

During the current reporting period, a total of 17 papers have been published as part of the REVEAL project. In addition to that, one brochure and one poster have been produced in order to be able to communicate the REVEAL project at events. Also, press releases have been produced and distributed by various partners.

The main dissemination channel and information hub for REVEAL activities is the project website (see http://revealproject.eu/). It has been decided to make this a "place to return to" by updating it frequently with both activities coming out of the project (to further increase in number as the project progresses and more research results emerge), as well as reporting about the field of social media verification and related activities in general. This is intended to have a reach also outside the research community, preparing for future exploitation. Furthermore, this approach has resulted in the REVEAL project as well as the EC who co-funds this undertaking obtaining quite some interest from people and organisations who have previously not been in touch with EC co-funded research. The figures / statistics show that we are on the right path and the dissemination strategy (as described in detail in deliverable D8.2.1) turned out to be very effective.

Finally, REVEAL has been a founding member of the European Center for Social Media (see http://www.socialmediacentre.eu for details). The aim of the initiative is to bring together social media research and researchers from across Europe and beyond.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 33 of 43

Task 8.2: Market analysis and exploitation planning

Similarly to previous task, this one started at project month M1 and will be active until M36. During the current reporting period the first version of the business exploitation plan was developed. This concerned the identification of the project's tangible assets which may be either exploited as individual open-source modules functioning on their own, as add-ons to existing commercial applications, or as an integrated solution offering a diverse range of services according to the use cases developed by tasks T1.1 and T1.2 of work-package one.

A market watch focusing to the latest trends in the social media with respect to the worldwide internet penetration, the worldwide IT spending forecast, the ICT emerging technologies and trends, the social media market evolution, and the digital news access and social media was conducted for evaluating the market opportunity for REVEAL; whereas a market analysis was performed in the news media and enterprise fields for delimiting the market sector for the project's exploitable assets, identifying the potential stakeholders interested to REVEAL outcomes, and listing the relevant and competitive products available in the market.

Taking into consideration the outcome of the market analysis along with the expertise and the initial individual exploitation intentions of each consortium partner, a preliminary study was made on how every REVEAL asset may be exploited at research and/or commercial level; and in a few cases the envisaged licensing scheme and associated IPRs were also provided. In addition, a SWOT analysis was prepared for the REVEAL foreseen products capitalizing upon the current project advances. It is important to state that the market watch and analysis will be a constant activity to run until the end of the project, and the exploitation plan will be enhanced and complemented as the project evolves.

The activities and outcomes of task 8.2 were documented in Deliverable D8.3.1 that was submitted in M12.

Task 8.3: Training activities

This task started three months earlier (at project month M9) than it was initially planned with the intention of exploiting the opportunity of providing a set of training activities under the umbrella of the well recognized International Research Centred Summer School (IRSS2014); held in Athens at NCSR'D premises from the 1st - 30th of July 2014. During the event, groups of undergraduate and postgraduate students were trained to technologies relevant to the objectives of the REVEAL project. In particular, issues related to contributors in social networks as well as to content were investigated by small groups of students under the supervision of NCSR'D researchers and participants to the REVEAL project.

Task 8.4: REVEAL community building

This task started one month before the end of the current reporting period (at project month M11). During this time, focus was placed on preparation and planning activities related to the definition of the community strategy, the design and establishment of the required infrastructure, as well as approaches on how the community can be created, maintained and exploited.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 34 of 43

Deliverables and milestone list for WP8

List of Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D8.1	REVEAL Website	8	DW	O	PU	M2 = Dec. 2013	Y	19 Dec. 2013
D8.2.1	Dissemination plan and initial set of activities	8	DW	R/P	PU	M8 = Jun. 2014	Y	30 Jun. 2014
D8.3.1	Exploitation plan	8	INTRA	R	PP	M12 = Oct. 2014	Y	31 Oct. 2014

List of Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
1	Project setup	1, 8, 9	INTRA	M6	Y	April 2014
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

4 Deliverables and milestones table

Below is a detailed list of all deliverables scheduled for submission during the first year of the REVEAL project along with their status (i.e. due date, delivery date etc.).

Table 1: REVEAL First Year (Y1) Deliverables

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Diss. Level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date
D1.1	Requirements analysis and specifications	1	ATC	R	PU	M6 = Apr. 2014	Y	5 May 2014
D1.2	Legal/regulatory requirements analysis	1	KU Leuven	R	PU	M6 = Apr. 2014	Y	5 May 2014
D2.1	Contributor modelling requirements and baseline approaches	2	NCSR'D'	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014
D3.1	Multimedia indexing requirements and baseline approaches	3	CERTH	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014
D4.1	Context extraction methods and context framework	4	UKob	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014
D5.1	Analytic report of metadata template definitions	5	ITINNO	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014
D6.1.1	Architecture and specifications	6	ATC	R	PU	M10 = Aug. 2014	Y	29 Aug. 2014
D8.1	REVEAL Website	8	DW	O	PU	M2 = Dec. 2013	Y	19 Dec. 2013
D8.2.1	Dissemination plan and initial set of activities	8	DW	R/P	PU	M8 = Jun. 2014	Y	30 Jun. 2014
D8.3.1	Exploitation plan	8	INTRA	R	PP	M12 = Oct. 2014	Y	31 Oct. 2014
D9.1	Project quality and assessment plan	9	INTRA	R	PP	M3 = Jan. 2014	Y	31 Jan. 2014
D9.2.1	Periodic activity and management reports	9	INTRA	R	CO	M12 = Oct. 2014	Y	17 Dec. 2014
D9.4.1	Public project reports	9	INTRA	R	PU	M12 = Oct. 2014	Y	17 Dec. 2014

The project work plan for the first year is formed around the following milestones which have been successfully achieved:

Table 2: REVEAL First Year (Y1) Milestones

Milestone No.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex I	Achieved Yes / No	Actual Forecast achievement date
1	Project setup	1, 8, 9	INTRA	M6	Y	April 2014
2	End of first period: Intermediate requirements and architecture. Preliminary research components	All	DW	M12	Y	October 2014

5 Annex

5.1 Dissemination events list

Table 3: List of performed dissemination events/actions within Y1

	Action	Target Group	Partner Responsible	Dissemination Activity Type	Means	Estimated Size of the Target Group	When	Expected Benefits
1	Press release on ATC's website	Journalists, general public, EC partners, academic	ATC	Press release	publication	Ca. 10.000	Dec. 2013	Raise awareness of the REVEAL project
2	Press release on ATC's LinkedIn company page	Journalists, general public, EC partners, academic	ATC	Press release	publication	Ca. 10.000	Dec. 2013	Raise awareness of the REVEAL project
3	Press release on ATC's LinkedIn group	Journalists, general public, EC partners, academic	ATC	Press release	publication	Ca. 10.000	Dec. 2013	Raise awareness of the REVEAL project
4	Press release on iLab website	Journalists, general public, EC partners, academic	ATC	Press release	publication	Ca. 10.000	Dec. 2013	Raise awareness of the REVEAL project
5	Presentation on "Linked Data Index Models for Context" at GESIS, Cologne	Research community	Uni Koblenz	Conference Presentation	Presentation	Ca. 50	06 Feb. 2014	Raise awareness of the REVEAL project
6	Presentation on "Linked Data Index Models for Context" at the 'Fachhochschule Rhein-Main', Wiesbaden	Research community	Uni Koblenz	Conference Presentation	Presentation	Ca. 50	24 Feb. 2014	Raise awareness of the REVEAL project
7	Presentation on "Verification of UGC: Challenges and Approaches" at a meeting for DW journalists and other interested parties at Deutsche Welle in Bonn & Berlin	Journalists, academic community	DW	Presentation	Presentation	Ca. 50	13 Mar. 2014	Raise awareness of the REVEAL project

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 37 of 43

	Action	Target Group	Partner Responsible	Dissemination Activity Type	Means	Estimated Size of the Target Group	When	Expected Benefits
8	Invited Talk on "Making Use of the Linked Data Cloud: The Role of Index Structures" at the 'Frühjahrstreffen der Fachgruppe Datenbanken'	Research community	Uni Koblenz	Talk	Presentation	Ca. 50	20 Mar. 2014	Raise awareness of the REVEAL project
9	Presentation at SNOW 2014, in conjunction with the 23rd International World Wide Web Conference (WWW2014) in Seoul, Korea	Research Community	ATC & DW	Conference Presentation	Presentation	Ca. 80	7-11 Apr. 2014	Raise awareness of the REVEAL project
10	Presentation at the 2nd International Symposium on Media Innovations, in Oslo, Norway	Research and industry community	SINTEF & DW	Conference presentation	Oral presentation	120	24-25, Apr. 2014	Presenting results and raise awareness of the REVAL-project
11	Presentation on "Contextual Data Retrieval" at GESIS, Cologne	Research community	Uni Koblenz	Conference Presentation	Presentation	Ca. 50	08 May 2014	Raise awareness of the REVEAL project
12	Presentation at the 26th International Conference on Advanced Information Systems Engineering (CAiSE 2014), in Thessaloniki, Greece	Research Community	ATC	Conference Presentation	Presentation	Ca. 50	16-20 Jun. 2014	Raise awareness of the REVEAL project
13	Presentation at the GeoSemantics Domain Working Group at the Open Geospatial Consortium (OGC) Technical Committee (TC) meeting in Geneva	Research Community	ITINNO	Conference Presentation	Presentation	200	20 Jun. 2014	Raise awareness of the REVEAL project

	Action	Target Group	Partner Responsible	Dissemination Activity Type	Means	Estimated Size of the Target Group	When	Expected Benefits
14	Poster presentation and participation in EC concertation meeting	EC ICT community	INTRASOFT & ATC	Poster presentation	Presentation	50	25 Jun. 2014	Raise awareness of the REVEAL project
15	REVEAL presentation at Deutsche Welle's Global Media Forum	Journalists,	DW	Conference Presentation	Presentation	200	30 Jun. - 2 Jul. 2014	Raise awareness of the REVEAL project
16	2014 IRSS Summer School	Students, research community	ATC & INTRASOFT	Presentation	Presentation	150	Jul. 2014	Raise awareness of the REVEAL project
17	Press release on "Luxembourg official", printed issue 23.9.2014	Journalists, general public	INTRASOFT	Press release	publication	10.000	23 Sep. 2014	Raise awareness of the REVEAL project
18	KDD-LESI Workshop	social scientists	NCSR Demokritos	Paper	Presentation	Ca. 50	24 Aug. 2014	Raise awareness of the REVEAL project
19	WebIT	Industry	Software AG	Demonstration	Presentation	10.000	1 Oct. 2014	Raise awareness of the REVEAL project

A list of scientific publications that were produced during the current reporting period is provided beneath:

1. "Influence Study on Hyper-Graphs", by Dimitrios Vogiatzis. Proceedings of the AAAI Fall Symposium 2013 on Social Networks and Social Contagion, Arlington, Nov 2013.

2. "Tweet Type, Location and Popularity: Case Study Hurricane Sandy", by Amanda Karavolia, Dimitrios Vogiatzis, Brandon Truong, Andrea Tapia, Cornelia Caragea, Anna Squicciarini and Georgios Paliouras. Paper accepted for KDD-LESI, 2014.

3. "Focused Exploration of Geospatial Context on Linked Open Data", by Gottron, Thomas; Schmitz, Johannes; Middleton, Stuart E. (2014). In: IESD'14: Workshop on Intelligent Exploration of Semantic Data.

4. "Intermediary Liability & Freedom of expression: Recent developments in the EU Notice & Action Initiative", by Aleksandra Kuczerawy. Paper accepted for the Computer Law and Security Review. It will be published in Vol 31.1 of CLSR.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 39 of 43

5. "A Unified Model for Socially Interconnected Multimedia-Enriched Objects", by Theodora Tsikrika, Katerina Andreadou, Anastasia Moumtzidou, Emmanouil Schinas, Symeon Papadopoulos, Stefanos Vrochidis, Yiannis Kompatsiaris. Paper accepted for 21th Multimedia Modelling Conference (MMM2015).

6. "Alethiometer: a Framework for Assessing Trustworthiness and Content Validity in Social Media", by Eva Jaho, Efstratios Tzoannos, Aris Papadopoulos and Nikos Sarris. Paper published in WWW'14 Companion, April 7–11, 2014, Seoul, Korea. ACM 978-1-4503-2745-9/14/04.

7. "News from the Crowd: Grassroots and Collaborative Journalism in the Digital Age", by Jochen Spangenberg and Nicolaus Heise. Paper published in WWW'14 Companion, April 7–11, 2014, Seoul, Korea. ACM 978-1-4503-2745-9/14/04.

8. "Social media as a trustworthy news source: Exploring journalists' working practices concerning social media" by Brandtzæg, P.B., Lüders, M., Spangenberg, J., Rath-Wiggins, L., & Følstad, A. (2014). Paper presented at The International Symposium on Media Innovations. April, 24-25, 2014. Oslo. [Received Conference Best Paper Award]. The paper was also presented in a research news outlet Forskning.no in Norway entitled "Can we trust news on Twitter?" on 26th of April, 2014. Retrieved from http://www.forskning.no/artikler/2014/april/388891

9. "Analysis of schema structures in the Linked Open Data graph based on unique subject URIs, pay-level domains, and vocabulary usage" by Thomas Gottron, Malte Knauf, Ansgar Scherp. In: Distributed and Parallel Databases, Springer, 2014, pp.1-39. 10.1007/s10619-014-7143-0

10. "One day in Twitter: Topic detection via Joint Complexity" by Gerard Burnside, Dimitrios Milioris, Philippe Jacquet. Proceedings of Snow Data Challenge, WWW'14 conference, pp.41-48. The paper achieved the 3rd prize/award of the SNOW data challenge.

11. "A Generalized Language Model as the Combination of Skipped n-grams and Modified-Kneser Ney Smoothing" by Rene Pickhardt, Thomas Gottron, Martin Körner, Paul Wagner, Till Speicher, Steffen Staab. ACL'14: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014 (forthcoming).

12. "Perplexity of Index Models over Evolving Linked Data" by Thomas Gottron, Christian Gottron. ESWC'14: Proceedings of the Extended Semantic Web Conference, Springer, 2014, pp. 161-175. 10.1007/978-3-319-07443-6_12.

13. "Providing Alternative Declarative Descriptions for Entity Sets using Parallel Concept Lattices" by Thomas Gottron, Ansgar Scherp, Stefan Scheglmann. ESWC'14: Proceedings of the Extended Semantic Web Conference, Springer, 2014, pp. 364-379. 10.1007/978-3-319-07443-6_31.

14. "Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling" by Johann Schaible, Thomas Gottron, Ansgar Scherp. At ESWC'14: Proceedings of the Extended Semantic Web Conference, Springer, 2014, pp. 457-472. 10.1007/978-3-319-07443-6_31.

15. "An Investigation of HTTP Header Information for Detecting Changes of Linked Open Data Sources" by Renata Dividino, Andre Kramer, Thomas Gottron. ESWC'14: Proceedings of the Extended Semantic Web Conference, 2014 (forthcoming).

16. "Of Sampling and Smoothing: Approximating Distributions over Linked Open Data" by Thomas Gottron. PROFILES'14: Proceedings of the Workshop on Dataset Profiling and Federated Search for Linked Data, 2014.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 40 of 43

17. "From Changes to Dynamics: Dynamics Analysis of Linked Open Data Sources" by Renata Dividino, Thomas Gottron, Ansgar Scherp, Gerd Gröner. PROFILES'14: Proceedings of the Workshop on Dataset Profiling and Federated Search for Linked Data, 2014.

Table 4: List of potential dissemination events and call for papers for Y2

Conference / Call for Papers	Date	Link
CeBIT 2015 – New Perspectives in IT Business	Mar. 16-20 2015	http://www.cebit.de/home
DIS 2015 – Digital Innovators Summit	Mar. 23-24 2015	http://www.innovators-summit.com/dis-home
IDEA/TOPOS – Social & Technological Innovation for Democracy	Mar. 25-28 2015	http://www.ideatopos.org
24th International World Wide Web Conference	May 18-22, 2015	http://www.www2015.it
Digital Media Europe	May 20-22 2015	http://www.wan-ifra.org/events/digital-media-europe-2015
ICWSM-15 (The 9th International AAAI Conference on weblogs and social media)	May 26-29, 2015	http://www.icwsm.org/2015/index.php
67th World Newspaper Congress and 22nd World Editors Forum	Jun. 1-3 2015	http://www.wan-ifra.org/events/67th-world-news-media-congress-22nd-world-editors-forum-25th-world-advertising-forum
Annual ACM International Conference on Multimedia Retrieval (ICMR 2015)	Jun, 23-26, 2015	http://www.icmr2015.org
IEEE International Conference on Multimedia and Expo (ICME 2015)	Jun. 29-July 3, 2015	http://www.icme2015.ieee-icme.org
21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining	Aug. 10-13, 2015	http://www.kdd.org/kdd2015
The 2015 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining	Aug. 25-28, 2015	http://asonam.cpsc.ucalgary.ca/2015
Software AG - Innovation World 2015	Oct. 12-14 2015	http://www.softwareag.com/IW15/index.php
CIKM 2015 (Conference on Information and Knowledge Management)	Oct. 19-23, 2015	http://www.cikmconference.org
Global Webit Congress 2015	Oct. 2015	http://webitcongress.com
ACM Conference On Online Social Networks (COSN'15)	*Not announced yet*	http://cosn.acm.org
SMWF 2015 - key event in the social & digital marketing conference	*Not announced yet*	http://www.socialmedia-forum.com/europe
GEN Summit 2015 – Global Editors Network	*Not announced yet*	http://www.globaleditorsnetwork.org

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 41 of 43

An additional list of relevant exhibitions and commercial events is provided within the Annex A section of Deliverable D8.3.1.

5.2   Report of the legal and ethical advisory board committee

This section provides a report of the first meeting of the REVEAL Ethical Advisory Board.

The legal and ethical issues presented by REVEAL are addressed in WP1, Task 1.3. This Task is led by the ICRI/CIR – KU Leuven. The role of this partner is to analyse the applicable legal framework, identify legal requirements and provide recommendations to the Consortium. Moreover, ICIR/CIR evaluates the project from a legal perspective and provides guidance towards legal compliance. The latter activity entails assisting partners in fulfilling their legal obligations (e.g., submitting notifications to national Data Protection Authorities or signing controller-processor and/or controller-controller agreements).

REVEAL is also advised by an external Ethical Advisory Board (EAB). This Board consists of three experts in the area of IT law, with special expertise relating to privacy and data protection. The members of the EAB are:

- Prof. Dr. Cecilia Magnussen-Sjoeberg, Swedish Law & Informatics Research Institute, Sweden;

- Prof. Dr. Nikolaus Forgo, Director of Institute for legal Informatics - IRI, Leibniz University of Hannover, Germany.

- Prof. Ioannis Iglezakis - Faculty of Law Aristotle University, Greece;

The members of the EAB met with the REVEAL project partners during the fourth General Assembly of the REVEAL project, which was held in Paris on 8-10 October 2014. During the meeting, the legal issues present in the project were discussed extensively. Moreover, the members of the EAB were provided additional information about the project in general and its specific aspects (e.g., user group requirements and technical components). All partners were represented during the meeting to answer any questions of the EAB members. Below is a short summary of the main topics discussed during the meeting.

In REVEAL three areas of law are crucial for the project, namely privacy and data protection, intermediary liability and media law. Additionally, the fundamental human right of freedom of expression is present in the contexts of all three of these areas. In the first year of the project the main focus was out on privacy and data protection. This issue was considered to be the most pressing with practical implications on the project. Since the project involves processing of personal data from social media, this aspect was also considered crucial from the perspective of ethical research and the privacy protection of individuals. The results of the conducted research were presented in deliverable D1.2 - Legal /regulatory requirements analysis. This document focused on the concepts and definitions as provided by the Data Protection Directive 95/46/EC[12] and its

---

[12] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281 , 23/11/1995 P. 0031 – 0050, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 42 of 43

transpositions into the Greek legal regime by the Data Protection Act (Law 2472/1997)[13]. This is because REVEAL will be implemented in Greece where the main data controller (ATC) is established. Deliverable D1.2 analysed also the proposed Data Protection Regulation[14] in order to prepare REVEAL for any changes that might be enacted during the project lifetime.

The main conclusion of the conducted research was that REVEAL must take necessary steps to ensure that processing of personal data in the project is ethical and in compliance with the applicable law. Processing will be based on two different legal grounds depending on the scenario: consent (enterprise scenario) and legitimate interest of the controller (journalistic scenario). The latter ground was analysed extensively in light of the recent opinion of the Art. 29 Working Party (Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC). Concepts such as 'reasonable expectation of privacy' in the context of REVEAL were discussed with the partners on numerous occasions. As a result a number of safeguards were developed that should ensure that privacy of social media users is not intruded upon. These safeguards refer to the principles of data minimisation, and transparency, as well as the right to objection and deletion. The proposed safeguards include provision of Privacy Policy of the REVEAL platform and its Terms & Conditions, as well as limitation of the obtained personal data to the minimum required to provide the service. Moreover, REVEAL will primarily target social media where users willingly participate in public discussions. Within those, only publically available profiles will be targeted. Another safeguard involves development of web forms that would allow individuals to object to processing of their data and request removal of content infringing their rights.

Other issue that was discussed was processing of sensitive personal data in REVEAL. The partners agreed that REVEAL will not intentionally target sensitive personal data (as defined by art. 8 of the Data Protection Directive). Incidental processing of sensitive personal data might occur. However, it would only concern data made manifestly public by the social media users. In such cases the prohibition of art. 8.1 DPD does not apply.

The conclusion of the meeting was that REVEAL partners have made good progress towards legal compliance. It was admitted by the EAB members that the legal issues present in the project are complex and not always (yet) addressed by the EU legislation. However, the attempt of the project partners to address them in an ethical and responsible way is clear. The EAB members recommended that in case of legal uncertainties a consultation with the Greek Data Protection Authority should be conducted (e.g., to examine the validity of a notification to the public in order to provide transparency). Moreover, the EAB members highlighted the relevance of the freedom of expression as a fundamental human right in REVEAL. It was suggested to place even stronger focus on this issue. Moreover, it was pointed out that due to the technological development the definition of journalism is constantly broadening. Attention to this interesting trend was recommended.

All the suggestions and recommendations of the EAB will be taken into account by the REVEAL project partners. They will be addressed in the following legal deliverables in Years 2 and 3 of the project. The progress in REVEAL will be further consulted with the EAB members.

---

[13] Law 2472/1997, http://www.dpa.gr/pls/portal/docs/PAGE/APDPX/ENGLISH_INDEX/LEGAL%20FRAMEWORK/LAW%202472-97-NOV2013-EN.PDF

[14] Proposal for a "REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data", http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.

Project Title: REVEAL   Contract No. FP7-610928
Project Coordinator: INTRASOFT International S.A.   **www.revealproject.eu**

Page 43 of 43