D2.2 Tracking technologies and social attitudes

Project acronym:	**SENSE4US**
Project full title:	**Data Insights for Policy Makers and Citizens**
Grant agreement no.:	**611242**
Responsible:	**Juri Papay**
Contributors:	**All Partners in the SENSE4US Consortium**
Document Reference:	**D2.2**
Dissemination Level:	**PU**
Version:	**FINAL**
Date:	**06/10/14**

Project co-funded by the European Union under the Seventh Framework Programme

History

Version	*Date*	*Modification reason*	*Modified by*
0.1	28/08/14	Initial draft	Juri Papay
0.2	08/09/14	Contributions from project partners	Juri Papay
0.3	29/09/14	Revision	Juri Papay
0.4	29/09/14	Draft reviewed by project partners	Hansard Society and GESIS
0.5	01/10/14	Final edits	Juri Papay
1.0	06/10/14	Submitted version	Juri Papay

Table of contents

List of tables

No table of figures entries found.

List of abbreviations

Abbreviation	*Description*
ALOE	Assisted Linked Data Consumption Engine
API	Application Programming Interface
DoW	Description of Work
LDA	Latent Dirichlet Allocation
LOD	Linked Open Data
Lemon	Lexicon Model for Ontologies
RDF	Resource Definition Framework
LOD	Linked Open Data
OR	Operational research
OWL	Web Ontology Language
PM	Project Month – PM1 is the first month of the project, etc.
PSM	Problem Structuring Methods
RDF	Resource Definition Framework
REST	Representational state transfer
SOFIE	Self-Organizing Framework for Information Extraction
SPARQL	Protocol and RDF Query Language
SVR	Support Vector Regression
WP	Work package

Executive summary

Sense4us is a three year project that was launched in October 2013 and co-funded under the Seventh Framework Programme (FP7-ICT-2013-10). Sense4us aims to assist policy makers by giving them tools to access a wide array of current data, take into account the views of citizens on policy issues in real time and help them to better understand the implications of proposed policies. Policy-makers at the EU, national (UK) and local (Germany) levels will be engaged in the development of Sense4us to ensure the toolkit has the widest possible application.

The main objective of the Sense4us project is to exploit the potential of open data and social networking sites for the benefit of policy makers and citizens. We expect that a wider access to information will also increase the participation of citizens in public matters and contribute to the transparency of the policy making process. One of the key elements of the Sense4us project is the integration of information from open data and social media. Although the data may refer to the same real world entity, the first provides objective facts while the other subjective thoughts and opinions. The integration of information originating from these data sources enables the composition of two complementary points of view of the same problem and thus provides a better, more in depth understanding of the subject area and can also support in the decision making process.

The Description of Work (DoW) specifies the requirements for Task 2.2 as follows:

> *Task 2.2 – Tracking other technologies and social attitudes (M03-33)*
>
> *In order to keep the project relevant and to grasp opportunities where they arise, research will be carried out and reports produced that take account of new technologies, social attitudes towards technologies and the impact of other research in the SENSE4US space. The outcome of this research will be fed into the decision making process of the project.*

In this deliverable we have surveyed the main directions of research, key concepts, technologies and social attitudes relevant for the Sense4us project. The outcome of this report will be used for the assessment of various options that arise during the functional and architecture design. In this first instalment of the deliverable we describe five main subject areas, these are:

- Document summarisation
- Finding related information
- Policy impact simulation
- Finding related public opinion
- Social attitudes

These subject areas cover the main components of Sense4us project and provide an overview of the alternatives which will be taken into account during the design. This is the first version of the deliverable, an update is due in month 24 that will summarise the lessons learnt during the project.

1 Introduction

Sense4us is a three year project that was launched in October 2013 and co-funded under the Seventh Framework Programme (FP7-ICT-2013-10). Sense4us aims to assist policy makers by giving them tools to access a wide array of current data, take into account the views of citizens on policy issues in real time and help them to better understand the implications of proposed policies. Policy-makers at the EU, national (UK) and local (Germany) levels will be engaged in the development of Sense4us to ensure the toolkit has the widest possible application.

One of the objectives of the Sense4us project is to exploit the potential of open data and social networking sites for the benefit of policy makers and citizens. We expect that a wider access to information will contribute to the policy making process by providing evidence for supporting the policies. One of the key themes of the Sense4us project is the integration of information from open data and social media. Although the data may refer to the same real world entity, the first provides objective facts while the other subjective thoughts and opinions. The integration of information originating from these data sources enables the composition of two complementary points of view of the same problem and thus provides a better, more in depth understanding of the subject area and can also support in the decision making process. Other objectives of the Sense4us project are: to advance the fields of policy modelling and simulation, data analytics and social network discussion dynamics, provide economic and social benefits to policy-makers on all governmental levels across Europe. These objectives in terms of software development can be translated into a set of tools that enable:

- the extraction of information from big data and open data sources;
- the automatic annotation and linkage of homogeneous data;
- the lexical analysis of sources;
- the creation of policy models combining quantitative open data sources with qualitative data from social media;
- the simulation of likely impacts of a policy; and
- the tracking of discussion dynamics in social media.

In this deliverable we'll review the state of art in the context of the above objectives according to the requirements for Task 2.2 specified in the Description of Work (DoW):

> *Task 2.2 – Tracking other technologies and social attitudes (M03-33)*
>
> *In order to keep the project relevant and to grasp opportunities where they arise, research will carried out and reports produced that take account of new technologies, social attitudes towards technologies and the impact of other research in the SENSE4US space. The outcome of this research will be fed into the decision making process of the project.*

In this deliverable we have surveyed the main directions of research, key concepts, technologies and social attitudes relevant for the Sense4us project. The outcome of this report will be used for the assessment of various options that arise during the functional and architecture design. In the first instalment of the deliverable we describe five main subject areas, these are:

- Document summarisation;
- Finding related information;
- Policy impact simulation;
- Finding related public opinion; and
- Social attitudes.

These subject areas cover the main components of the project and give an overview of the alternatives which will be taken into account during the design. This is the first version of the deliverable, the update is due in month 24 that will summarise the lessons learnt during the project.

The document is organised according to the identified subject areas. Section 2 describes document summarisation. The issues related to Open Data and information search can be found in Section 3. Section 4 details policy impact simulation and Section 5 outlines techniques for discovering public opinion. Section 6 discusses social attitudes in relation to politics, in general, and to the policy making process specifically. Section 7 summarises the findings of the report and makes recommendations for the Sense4us project.

2 Document summarisation

2.1 Definition

The purpose of document summarization is to reduce a large volume of text to a summary that contains the key points of the original document. The produced summary must be both concise and comprehensive. The key topics contained in the document (or in multiple documents) should be clearly identified and discussed from different perspectives. One of the advantages of automatic document summarisation is that it can produce an output without human intervention and therefore it is unbiased. However this output must be assessed for accuracy and consistency before it is published (Mani 2014).

In general there are two techniques used for producing a summary, these are: extraction and abstraction (Mani 2014). The extractive approach selects a subset of existing words, phrases, or sentences in the original text for generating a summary. The abstractive approach builds a semantic representation of the text and then uses natural language generation techniques for creating a summary. In this case the produced summary might contain words that are not explicitly present in the original text. Abstractive methods represent a promising direction of research, but the majority of summarisation tools still use extractive methods.

There are several issues related to document summarisation, these have been derived from several publications (Mani 2014), (Lloret 2010), (Nenkova 2011):

a) dealing with the thematic diversity of a large number of documents;
b) combining the main themes with completeness;
c) readability;
d) conciseness;
e) representing a wide diversity of views on the topic;
f) reflecting the logical structure of the main content;
g) dividing the output into meaningful paragraphs;
h) transition from general topics to more specific themes; and
i) consistent referencing of information sources.

The detailed analysis of the above listed issues is beyond the scope of this report and we'll focus only on the themes that have been identified by the research partners and considered to be relevant in the context of the Sense4us project. These themes are:

a) automatic labelling;
b) topic extraction;
c) semantic mapping; and
d) semantic search.

In the following sections provide a short survey of these themes and describe the current state of art.

2.2 Automatic labelling

Labelling is the first step for determining the topics that a document addresses. The most generic approach to automatic labelling uses primitive labels as described in (Griffiths 2004),

(Blei, D. M, et al 2003). In this technique the top words are ranked using the probabilities associated with a given topic.

Current research however, indicates that selecting the top terms is not sufficient for interpreting the coherent meaning of a given topic (Mei, Q., et al 2007). More recent approaches have explored the use of external sources for example DBpedia and Wikipedia for supporting automatic labelling of topics by deriving so called *candidate labels* by using lexical (Lau, J.H., et al 2011), or graph-based (Hulpus 2013) algorithms. The paper authored by Mei *et al* proposed an unsupervised probabilistic algorithm for automatic assigning of labels in a topic model (Mei, Q., et al 2007). The proposed approach was defined as an optimisation problem involving the minimisation of the Kullback–Leibler divergence (Kullback-Leiber 2014) between the given topic and the candidate labels while maximising the mutual information between these two words. It was proposed by (Lau, J.H., et al 2011) to label topics by selecting the "top n" terms by using different ranking mechanisms including point-wise mutual information and conditional probabilities.

Another frequently used technique for automatic labelling utilises external data sources such as dictionaries, thesauruses or specialised topic ontologies. Methods relying on external sources for automatic labelling have been described by Magatti *et al* (Magatti,D., et al 2009). The authors derived candidate topic labels using the Latent Dirichlet Allocation (LDA) algorithm (LDA 2014) and the hierarchy obtained from the Google Directory service. Then the initial candidate labels were extended with the Open Office English Thesaurus. In their paper, (Lau, J.H., et al 2011), described a case study where the label candidates for topics were extracted from Wikipedia articles.

One of the promising approaches in this area is the construction of a Support Vector Regression (SVR) model that enables the ranking of label candidates (Support 2014). More recently, (Hulpus 2013) proposed the use of structured data sources (for example DBpedia) and graph centrality measures for generating semantic concept labels.

2.3 Topic extraction

Topic extraction is a starting point for various tasks, for example text summarisation and sentiment analysis. One of the main challenges is understanding the meaning of topics by ensuring "semantic coherence" (Aletras 2013). Recently numerous tools have been developed to enable topic extraction. These tools usually provide REST API and also endpoints for Standard Development Kits supporting popular programming languages. In this section we survey the following tools:

a) Semantria's (Semantria 2014);
b) Texalytics (Texalytics 2014);
c) AlchemyAPI (AlchemyAPI 2014); and
d) Yahoo Content Analysis API (YahooAPI 2014).

Semantria (Semantria 2014) allows topics to be extracted that are relevant for single or a collection of documents. Semantria relies on Wikipedia's ontology for constructing a concept matrix by using a deep learning algorithm. The tool offers a simple API that allows the formulation of queries, tagging, grouping and structuring the collection of derived topics.

Semantria also provides sentiment analysis for the data extracted from the social media. The API can be accessed directly from Excel via a plug-in. One of the big benefits of this tool is the customisation of output by editing the categories and tags in the configuration file.

Topic extraction in **Textalytics** (Texalytics 2014) is based on combining a number of complex natural language processing techniques that produce morphological, syntactic and semantic analysis. The tool produces different types of elements which are classified as:

- Named entities: people, organizations, places, etc;
- Concepts: significant keywords in the text;
- Time expressions;
- Money expressions;
- URIs;
- Phone number expressions;
- Other expressions: alphanumeric patterns;
- Quotations; and
- Relationships.

The **AlchemyAPI** (AlchemyAPI 2014) tool provides topic extraction, image tagging and sentiment analysis of web pages, documents, tweets and photos. AlchemyAPI uses neural network algorithms for the linguistic and statistical processing of large volumes of text. This tool is frequently is used in the business environment for extracting knowledge about the customers, competitors, company operations, marketing, sales and new product developments. AlchemyAPI provides language support for Python, PHP, Ruby, Java, Perl, C/C++, C#(.NET) and Android OS.

The **Yahoo Content Analysis API** (YahooAPI 2014) is a Web Service that allows the detection of entities/concepts, categories, and relationships in the text. The service ranks the detected entities and concepts by their overall relevance. These items are then linked to Wikipedia pages and annotated with tags and relevant meta-data.

2.4 Semantic mapping

Semantic mapping can be described as a transformation of data from one namespace into another (Suchanek F. M. 2009). In the context of the Sense4us project the data is represented by topics which are mapped into semantic relationships. These relationships can be expressed by semantic nets or ontologies. The task in this case can be described as a three stage process: extracting the topics from the text, creating a semantic network of topics and linking the topics to Open Data resources. In this section we survey three tools that might be relevant for Sense4us due to the fact that their functionality a close match with the original aims of the project. These tools are:

- LODifier (LODifier 2014);
- SOFIE (Self-Organizing Framework for Information Extraction) (Suchanek F. M. 2009); and

- Lemon (Lexicon Model for Ontologies) (McCrae 2011).

The main purpose of LODifier (as the name indicates) is converting unstructured natural language into Linked Data (LODifier 2014). LODifier uses deep semantic analysis, word-sense disambiguation, Semantic Web vocabularies for extracting entities and relatioships between them. The output is translated into RDF representation where the individual nodes are linked to DBpedia and WordNet. The key features of LODifier are: abstracting away from linguistic surface variations, using a semantic graph for representing explicit structural information and linking up the concepts and relations to the LOD cloud (for additional details see (Augenstein, I et al 2012).

The SOFIE system enables the integration of the discovered data into existing knowledge bases (Suchanek F. M. 2009). SOFIE can parse documents, extract ontological facts and link the discovered facts into ontologies. Algorithms based on logical reasoning are used for identifying the most probable meaning of words. This process relies on searching for text patterns in a knowledge database. The key idea of SOPHIE is that the knowledge stored in an existing ontology can be used for gathering new data and also for reasoning about hypotheses. This feature enables not only a growth but also validation and restructuring of an ontology.

Lemon (Lexicon Model for Ontologies) is a higher level model that enables the linking of lexical resources to ontologies and thereby providing input for Natural Language Processing (NLP) (McCrae 2011). The paper also described that the existing RDF model of WordNet (WordNet 2014) can be mapped into Lemon's semantic network and thus enabling linking with other ontologies.

2.5 Semantic search

The purpose of semantic search is to use the conceptual meaning of terms and also the intention of the user for improving the accuracy of output. Semantic search takes into account various aspects, for example, context of search, location, intent, synonyms of words, natural language queries, concept matching etc. for providing relevant search results (John 2012). The popular search engines such as Bing and Google also use elements of semantic search for finding the most relevant information. In the literature we can find references to five techniques used in semantic search (Makela 2005), these are:

- RDF Path Traversal;
- Keyword to Concept Mapping;
- Graph Patterns;
- Logics - by using inference based on OWL; and
- Fuzzy concepts, relations and logics.

2.6 Document summarisation toolkits

There are several text summarisation toolkits used in practice that represent interest for the Sense4us project. These are the Ultimate Research Assistant (Assistant 2014) and iResearch

Reporter (iResearch 2014). There are also tools specialised for processing news items, these are: Newsblaster (Newsblaster 2014), NewsFeed Researcher (Researcher 2014), NewsInEssence (Radev 2005). Typical features of these tools are: text mining of documents, concept extraction, hierarchical concept clustering, text summarisation, taxonomy generation and various visualisation tools (for example mind map generation).

3 Finding related information

3.1 Definition

There are several data resources considered by the Sense4us project, these are: Open Data, social media, academic research, policy-makers' own data and policy documents. The overall objective is to extract information that is relevant for a specific subject area. In the context of the Sense4us project Linked Open Data (LOD) is the main source of information. The original concept of Linked Open Data was introduced by (Berners-Lee 2006). This concept was derived from the Semantic Web (Berners-Lee, T. et al 2001), which proposed an enhancement of the Web and the resources published on the Web with more expressive semantics. The authors suggested that this semantic augmentation could be achieved by establishing and describing hyperlinks between Web resources. Heath and Bizer (Heath 2011) in their paper summarised the benefits of Linked Data as follows:

> "*Linked Data provides a unifying data model a standardized data access mechanism, hyperlink-based data discovery and self-descriptive data*".

In the context of the Sense4us project based on the communication with the project partners we can identify four main themes relevant for the usage of data resources, these are:

a) integration of heterogeneous data sources;
b) data linking;
c) information discovery; and
d) data visualisation.

These themes will be described in details in the following sections.

3.2 Data integration

The term "*data integration*" refers to the problem of combining data from different sources and providing the user with a unified view of this data (Lenzerini 2002). The main issues in this respect are the different structures, different formats and different dynamics of data. With the growing volumes of data available on the Web, it has been recognised that the traditional data integration approaches, which assumed a limited number of relational data sources, are inadequate for the challenges of the Web. This led to the emergence of so called "*dataspaces"* which represent an abstraction for data integration. The generic vision of a dataspace is based on the following principles (Franklin, M., et al 2005):

a) providing support for a wide variety of data formats;
b) offering various levels of service depending on the type of query and parameters of data sources; and
c) integration of data sources.

A good survey of dataspace solutions can be found in (Hedeler, C., et al 2010). In their paper, (Cafarella, M., et al 2011), described two Google projects: WebTables and Google Deep Web Crawler, which investigated the problem of structuring information published on the web. The WebTables project compiled a large collection of databases by crawling the Web and trying to combine the tables that are embedded in HTML pages. The Google Deep Web Crawler attempted to extract data from the so called Deep Web that is made up by a large

number of Web forms. These two projects tried to harvest the vast amount of data on the Web that cannot be found by the traditional web crawlers.

There are numerous approaches to data integration, in respect to the Sense4us project the following four systems might be relevant:

a) Sindice + Sig.ma (Tummarello, G., et al 2010) aim to provide light-weight keyword-based search over semantic data collected from the Web;
b) Information Workbench (Haase 2011) is intended as a platform to support Linked Data application development. The system primarily targets the virtual integration of data sources, which are accessed using query federation;
c) LDIF (Schulte, A., et al 2011) provides a suite of tools for the integration of data retrieved from the Web. Important features of the system are data translation and co-reference resolution; and
d) ALOE (Assisted Linked Data Consumption Engine) (ALOE 2014) provides a linked data consumption engine. Consumption of Linked Data is one of the relevant tasks in relation to the increasingly complex Web of Data. This complexity is due to the rapidly growing size of data, the diversity of vocabularies that describe the data and the lack of a common schema for knowledge bases. ALOE provides a semi-automatic solution to address this problem by using light-weight techniques for generating schema and data-level mappings.

3.3 Data linking

The objective of data linking is to establish hyperlinks between related data items that are stored in different data sources. Once the data is linked it can be accessed by using Semantic Web browsers in a similar way to the traditional web browsers. The difference is however, that in the case of the semantic search the users follow RDF links for navigating between data sources. The RDF links can also be followed automatically by specialised robots and semantic search engines (Heath 2011).

Publishing data on the web is as easy as publishing HTML pages. However there are general recommendations described in Tim Berners-Lee's paper that allow the data to be found (Berners-Lee 2006):

a) use URIs as names for things;
b) use HTTP URIs so that people can look up those names;
c) provide useful information, using the standards (RDF, SPARQL); and
d) include links to other URIs.

In relation to data linking there are several initiatives that Sense4us can benefit from, these are Microformats (Microformats 2014) and Linked Data (LinkedData 2014). The aim of Microformats (Microformats 2014) is to extend the web with structured data. The idea is to define a set of simple data formats that can be embedded into HTML via class attributes. The data items that are included in HTML pages via Microformats do not have their own identifier. This makes linking the data across documents and Web sites a major issue. The Linked Data initiative promotes a set of best practices for publishing and connecting structured data on the Web. The key technologies used by Linked Data are URI, HTTP and RDF.

3.4 Information discovery

Linked Data is typically published either as static RDF files on a web server, embedded into HTML files or via database access (relational databases or RDF triple stores). An essential requirement for the publication of Linked Data is that the URIs of resources can be looked up by HTTP clients. Linked Data can be consumed in different ways. It can be crawled and integrated by specialised web-spiders for providing a consistent view of the data (see LDSpider (Isele 2011)). Data can be consumed also by de-referencing only the URIs that are currently in use by the given application. Linked Data applications using this method are typically Linked Data Browsers. Techniques for on-the-fly dereferencing are presented in Hartig et al., (Hartig, O., et al 2009).

Some applications enable to federate the query process by sending complex SPARQL queries SPARQL endpoints which are identified in advance, see (Langegger 2008). Link Discovery Tools are mainly similarity-based but they also apply machine-learning techniques. Tools that we consider to be relevant in the context of the Sense4us project are the following:

a) Silk (Volz, J. et al 2009);
b) LIMES (Ngomo, A. et al 2011);
c) SERIMI (Araújo 2011);
d) Amalgame (van Ossenbruggen 2011); and
e) RiMOM (Li, J. et al 2009).

Another approach for improving information discovery is based the creation of source catalogues and meta-level descriptions that summarise the information stored in the repositories. Examples of this concept are (CKAN 2014), (VoId 2014) and (DCAT 2014) catalogues.

3.5 Data visualisation

A suitable presentation of linked data is one of the critical requirements as described in (Halevy 2012). The visualisation is usually based on the hierarchical graph structure that allows not only easy navigation but also better understanding and summarising of large amounts of data (Wills 2009), (Ell 2011). In the context of the Sense4us project we recommend taking a closer look at the following initiatives: LESS (Auer 2010), Fusion (Araujo 2010) and TWC portal (Ding 2011).

The LESS framework enables end-to-end integration of linked data from multiple sources based on parameterised data and queries (Auer 2010). LESS also uses a template mechanism which allows flexible integration of different formats, such as text, plots, graphs, maps, etc. The visualisation used in the Fusion project (Araujo 2010) allows mappings between the source ontology of linked data and the application ontology for the processing and visual presentation of information. Various demos of linked data visualization have been implemented in the TWC portal which demonstrates examples of linked governmental data (Ding 2011). The presentation layer of the TWC LOGD portal (Tetherless World Constellation (TWC) and Linking Open Government Data (LOGD)) uses mash-ups based on the Google

visualization API. In the Fusion Tables project Google also provides a framework for data integration and visualization (Gonzalez 2010).

4 Policy impact simulation

4.1 Definition

The academic literature describes decision making as a process of identifying and choosing alternatives based on the considerations and preferences of the decision maker. The objective of this process is to reduce the number of alternatives so that a favourable outcome can be selected. This is an iterative process which requires the continuous checking of initial assumptions, parameters and outcomes (Harris 2012). Decision support tools and techniques enable the policy maker to predict the consequences of a policy before they are enacted. With these tools the user can carry out "*what if"* studies that allow a given policy to be analysed from the perspective of different stakeholders and for the likely effects on society and groups of individuals (Jiwani 2010), (Mitchell 2009).

One of the key components of a decision support system is *modelling* which incorporates several activities. In relation to modelling the following activities can be identified (Druckenmiller et al 2009), (Mintzberg 1994):

a) describing the scenarios;
b) identifying the key variables, ranges of variables and uncertainties;
c) describing the stakeholders and preferences;
d) specifying the assumptions of the model;
e) specifying the mathematical model or a simulation tool;
f) interpreting of the output; and
g) providing feedback to the modelling process.

4.2 Modelling techniques

In the literature we can find various approaches to modelling, an overview of various techniques can be found in (Mingers 2004). In the context of Sense4us project the following techniques have been considered:

a) Operational research and Problem Structuring Methods (PSM);
b) Multi-Criteria Decision Analysis;
c) Multi-attribute utility theory;
d) Portfolio decision analysis; and
e) Game theory.

Operational Research (OR) is one of the well-established methodologies used for modelling and decision support. There are numerous tools developed in OR that can be used for the evaluation of alternative scenarios, see details in (Weimer and Vining 2005), (Gass 2011), (OR 2014). There are also advanced techniques that aim to extend the traditional OR methods by taking into account the qualitative data in the modelling process. These techniques are referred to as Problem Structuring Methods (PSM) (Mingers 2004). The traditional OR mainly works with well-structured problems, the PSM on the other hand tackles problems which are unstructured and contain many uncertainties. Typical features of unstructured problems are multiple actors, multiple perspectives, conflicting interests, important intangibles and large number of uncertainties (Mingers 2004). A detailed account of PSM is beyond the scope of

this report, therefore we just mention the typical examples of PSM techniques (J. Rosenhead 1989):

- Strategic options development and analysis (SODA) (Eden 2000);
- Soft systems methodology (SSM) (Connell 2001);
- Strategic choice approach (SCA) (Friend 2001);
- Drama theory (Rosenhead 1996);
- Viable systems model (VSM) (Harnden 1990); and
- Decision conferencing (Phillips 1989).

The information available for modelling can be either quantitative or qualitative. The quantitative information can be integrated into mathematical expressions while the qualitative information is essential for the conceptualisation and formulation of the model (Sterman 2000). Needless to say it is harder to deal with qualitative data since there are generally no accepted rules for obtaining, analysis and interpreting this type of information. The problems involved with qualitative information can be summarised by the following quotation from (Luna-Reyes et al 2003):

"Although there is general agreement about the importance of qualitative data during the development of a system dynamics model, there is not a clear description about how or when to use it. The lack of an integrated set of procedures to obtain and analyze qualitative information creates, among several possible problems, a gap between the problem modeled and the model of the problem."

Another approach to modelling is based on Multi-Criteria Decision Analysis (MCDA). This method is widely used in complex decision problems for evaluating various alternatives leading to the achievement of multiple objectives (Keeney R.L 1993). One of the assumptions of this technique is that the alternatives are defined at the beginning of the modelling process and they do not change. However, in practice these alternatives are often difficult to specify and they can also change as the modelling work progresses and more information becomes available (Franco L. A. 2011).

A recently developed method based Multi-attribute Utility Theory proposes an alternative approach for integrating both quantitative and qualitative data into the same model (Danielson, M., et al 2006). The main idea of Portfolio Decision Analysis (PDA) is based on the selection of a portfolio of alternatives (instead of a single one only) which satisfies certain resource constraints, for example costs, budget etc. (Fasth 2012).

Game theory in general terms can be described as the "*the study of mathematical models of conflict and cooperation between intelligent rational decision-makers*" (Myerson 1991). This theory is used in numerous fields such as economics, politics, management, defence, biology etc. This theory assumes that games are well defined in terms of players, rules and outcomes. The objective is to design game strategies that can produce the desired outcome.

4.3 Modelling Tools

There are numerous design tools that allow conceptual and simulation models to be created for assessing the consequences of various policy decisions. According to this survey of these tools, the following categories can be identified (tools 2014):

- Information Control - gathering, storage, retrieval, and organisation of data, information and knowledge;
- Paradigm Models - paradigms, frameworks or perspectives for conceptualising the model;
- Simulation Models - models that provide answers to "*what if*" questions;
- Ways of Choosing - reducing the number of alternatives; and
- Representation Aids - visualisation of data or problem space.

In the context of the case study ("green energy") investigated by the Sense4us project we can also highlight an award winning tool "2050 carbon calculator" developed by the *UK Department for Energy and Climate Change.* The tool allows users to simulate the outcomes of different energy policy scenarios, and the impact of their choices on the climate (calculator 2014). This climate calculator produces emission reduction pathways and demonstrates the impact of real scientific data. The tool also allows a user to check whether the data included in policy documents supports the long term objectives on emissions reduction. A further example of a similar tool, extended for the entire world is called the *Global Calculator* which allows a user to estimate the impact of the world's energy, land and food systems on the climate (Global 2014).

5 Finding related public opinion

5.1 Information from social media

The objective of social media analysis in the context of Sense4us is to collect, analyse, track and interpret the discussions and the exchange of information related to specific topics. The information found in social media has certainly got the potential to improve the quality of evidence required for policy document drafting (Leavy 2013). There are however, several concerns about using this source of information in the context of policy making or for assessing the public reaction to various policies. One of the concerns is the lack of information about the participants of political discussions on the social media (Wheeler 2012). The recent data collected by (Fernandez, M. et al 2012) show that only a small percentage of users generate most of the discussions, in real terms about 6% generate more than 36% of all collected tweets. It is also important to mention that most of these tweets are generated by professional "opinion forming" organisations such as news agencies and think tanks rather than individual citizens. This finding, although it is only a preliminary study, raises many questions regarding the use of information coming from the social media for assessing public opinion.

5.2 Sentiment analysis

The objective of sentiment analysis is to extract evidence from social media that can support the policy making process and to get feedback on certain topics. This involves the monitoring of public debate, analysing the dynamics of discussions and determining public opinion in regard to specific policies. Data mining of social media, along with the development of mechanisms and tools for harnessing this rich content are the subject of intensive research. The main directions of this research have been outlined by (Agarwal, A., et al 2011) and (Beevolve 2012):

a) collecting data on the evolution of political debates;
b) monitoring the change of sentiment;
c) modelling the evolution of a political debate;
d) providing information from open data sources to enrich the political debate; and
e) extracting evidence from debates and summarising the key arguments.

Recent studies have shown a strong correlation between the online opinion of individuals and their offline attitude towards political issues in the real world (Hussain 2011). Hence tracking public opinion on various social media platforms can provide important insights for policy makers. By collecting information we can get a picture of the citizens' attitudes towards policies as well as positive and negative arguments. According to The IBM Center for The Business of Government "*Next Four Years: Citizen Participation*" more and more people are turning to social media to discuss their political views (Citizen Participation 2012). Today, some social networking sites have managed to attract millions of active users, and about 20% of all Internet traffic (Traffic 2014). Social media have therefore become a rich source of content and an easy and quick way to reach out to millions of people. While social networks

are known to be highly accepted among younger generations, they are also gaining popularity with older generations (Madden 2014).

Most existing approaches to sentiment analysis focus on classifying the individual words into subjective (positive or negative) or objective (neutral). They can be categorised by supervised approaches and lexicon-based techniques. Supervised methods are based on training classifiers from various combinations of features such as word "*n-grams*" (Pak 2010), Part-Of-Speech (POS) tags (Agarwal, A., et al 2011) and tweets syntax features such as hash-tags, re-tweets, punctuations, etc. (Kouloumpis 2011). These methods can achieve around 80% accuracy, however, training data is usually difficult to obtain especially for continuously evolving subject domains for example in Twitter (Liu 2010).

Lexicon-based methods use a dictionary of opinionated words for example; great, sad, excellent etc. for calculating the overall sentiment (Bollen 2011). Examples of methods based on sentiment dictionaries are SentiWordNet (sentiwordnet 2014), MPQA subjectivity lexicon (Wilson 2005) and SentiStrength (sentistrength 2014).

Another approach to sentiment analysis is based on ontologies and natural language processing techniques capturing the conceptual representations of words (Saif, H. et al 2012). The results produced by conceptual semantic approaches are better than the output produced by syntactical approaches, however they are often limited by the size and scope of knowledge bases (Cambria 2013).

Sentiment analysis was one of the key components of two previous projects WeGov (WeGov 2014) and ROBUST (ROBUST 2014). In the WeGov project (WeGov 2014) several generic discussion analysis methods were implemented. These methods allowed the analysis of the type of content and social features and also identified the behavioural roles of participants. In the ROBUST project (ROBUST 2014) the behavioural models were extended by applying these models to larger communities. In both WeGov and ROBUST projects, the analysis was focused on capturing and understanding the state of a discussion at any given point in time.

6 Social attitudes

6.1 Definition

In this section we give a short overview of the social attitudes of the public to policies and to the policy making process in general. We also discuss the ways that the public can engage and contribute to policy making by participating in debates, providing feedback and suggestions to decision makers. The term, "*social attitude*", is a wide concept and perhaps can be captured by the following quotation:

"Social attitude is a person or groups reaction to other people, races, cultures, ideas, or traits. Social attitude measures the person or groups like or dislike toward a certain subject."
(Answers 2014)

The heterogeneity of individuals and groups within society will naturally result in different reactions to policies. Some policies may change the behaviour and preferences of social groups or individuals.

6.2 Engaging the public in policy making

The participation of the public in politics is regularly measured and evaluated. According to statistics (Politics 2014) the overall interest of public the in politics is low. In the context of the UK the following quotation can illustrate this point.

"There is a common concern that the British public is increasingly becoming disengaged with politics." (Politics 2014)

In terms of policy making it is important to consider the behavioural characteristics of social groups and individuals. Regarding behaviour, the following observations can be made (Politics 2014):

- people display '*bounded rationality*' since they do not have access to the data and have not got the time for complex calculations and reasoning;
- people's preferences change over time – it also often happens that short term preferences are not in line with long term goals;
- people exhibit reciprocity and value fairness.

For engaging the public in policy related debates according to DEFRA's study the following so called "*4E*" requirements should be addressed (Collier 2010):

- Encouraging - giving the right signals, incentives and disincentives ensuring that the target audience responds;
- Enabling - putting in place the capacity, infrastructure, services, skills, guidance, information and support needed;
- Engaging - getting people involved; and

- Exemplifying - leading by example and sharing responsibility.

6.3 Public perceptions of the policy making process

The latest figures from the 2014 Hansard Society report into political engagement highlight the following (Audit 2014):

- 50% of the British public say they are *'very'* or *'fairly'* interested in politics;
- 50% of the British public say they feel they know *'a great deal*' or *'a fair amount*' about politics;
- 49% of the British public say they are certain to vote in the event of an immediate general election; and
- 33% of the British public think that the system of governing in this country works well.

Despite the relatively high level of interest in politics the policy making process itself to a large extent has been disconnected from the public. In other words there is little if any input from the public into the policy making process. It is also difficult to measure the feedback from the public regarding the policies that have already been implemented. As one of the benefits of the rapid development of social media, it is expected that the public will have a greater say in the policy making process. However, there are still barriers that prevent policy makers from becoming engaged with the citizens in political debates. These aspects can be summarised as follows (Collier 2010):

a) citizens feel that they are not well enough informed about the way decisions are made;
b) they also think that they do not have much influence over decision making;
c) most of the information which the public receive and react to is delivered via traditional media such as television and news;
d) the news items are interpreted by professional journalists and often published with editorial bias; and
e) the supporting evidence and data are not presented or are difficult to access. This makes the validation of policies difficult if not impossible.

7 Summary

In this deliverable we have surveyed the main directions of research, the key concepts and tools that we consider to be relevant for the Sense4us project. The project itself covers a wide range of topics that involve open data sources, harvesting information from social media, the visualisation of semantic data, and the analysis of social and demographics aspects in policy making. Producing a comprehensive survey of all these areas is far beyond the capacity and scope of the presented report. Therefore we have focused only on the themes that cover the key components of the Sense4us system. The themes surveyed in this report included: document summarisation, finding related information, policy impact simulation, finding related public opinion and social attitudes. In each of these subject areas we have provided numerous references to the literature that capture the main directions of research, the key concepts and tools.

The intention of this report was to highlight the opportunities, take account of new technologies and provide pointers to alternatives that can be used for the functional specification and architecture design. This is the first instalment of the deliverable, the updated version will be published in month 24, which will summarise the lessons learnt during the project.

The findings of the report for individual subject areas can be summarised as follows:

a) *Document summarisation,* in light of the rapidly increasing volume of digital publishing, is becoming one of the key research areas of computer science. This research promises the automatic production of digests that are concise, comprehensive and also reflect the key points of input documents. Recently there has been a significant progress in this direction, however there are only a few studies that provide a critical assessment of the quality of output against the digests produced by humans.
b) *Finding related information* that can aid reasoning and decision making is essential for drafting policies. If we looked at the information that is published on the web, it is mainly text based. Although there is an increasing amount of data, it is mainly hidden in databases and difficult to find since the data is not indexed by the search engines. The Linked Open Data initiative is trying to improve this situation by integrating data sources and making them public. However, it must be said that until data search becomes an integral part of standard web browsers the wealth of data will remain hidden.
c) *Policy impact simulation* aims to develop analytical and simulation models that allow the investigation of various "*what if*" studies by changing the numerical input values of the model. The main issue in this respect is interpreting the policy document(s) in terms of a numerical simulation model. This is a complex task and it is unlikely that it can be achieved automatically. Although some rudimentary model can probably be extracted directly from the text, further conceptualisation, refinement and validation of this model most certainly will require human intervention.
d) *Finding related public information* in social media has got the potential to gauge the mood of the society in relation to various events and policies. However, it must also

be said that this information can be distorted since according to recent studies most of the traffic on certain topics is generated by a small fraction of participants. One of the main problems with harvesting social media is the lack information about the profile of participants which makes any sound statistical studies about the real public opinion difficult to conduct.

e) *Social attitudes to politics* are regularly monitored and measured by numerous organisations. The reason is to measure the level of public engagement and also to find out whether there is acceptance of a policy. Although some citizens are interested in politics they have little if any involvement in the process of policy making itself. There are only a few opportunities (if any) to express opinions or to provide input to the policy making process. The consequence is that there is a lack of understanding why certain policies have been accepted and what was the justification for their introduction.

8 References

Aday, S. *Blogs and bullets: New media in contentious politics.* 2010. http://www.usip.org/files/resources/pw65.pdf.

Agarwal, A., et al. "Sentiment analysis of twitter data." *In proceedings of the acl 2011 workshop on languages in social media .* 2011. 30–38.

AlchemyAPI. 2014. http://www.alchemyapi.com/) .

Aletras, N. and Stevenson, M. " Evaluating topic coherence using distributional semantics on Computational Semantics (IWCS 2013)." *In Proceedings of the 10th International Conference .* Potsdam: Association for Computational Linguistics, 2013. 13–22.

ALOE. 2014. http://aksw.org/projects/aloe.

Answers. 2014. http://www.answers.com/Q/What_is_the_definition_of_social_attitude.

Araujo, S. et al. "Fusion – visually exploring and eliciting relationships in linked data." *The Semantic Web – ISWC 2010.* Springer Berlin / Heidelberg, 2010. 1–15.

Araújo, S., et al. "SERIMI - resource description similarity, RDF instance matching and interlinking." *Proceedings of the 7th International Workshop on Ontology Matching.* 2011.

Assistant, Ultimate Research. 2014. http://www.ultimate-research-assistant.com/GenerateResearchReport.asp.

Audit. 2014. http://www.auditofpoliticalengagement.org/.

Auer, S. et al. "Less template-based syndication and presentation of linked data." *The Semantic Web: Research and Applications.* Springer Berlin / Heidelberg, 2010. 211–224.

Augenstein, I et al. "LODifier: generating linked data from unstructured text." *ESWC'12 Proceedings of the 9th international conference on The Semantic Web: research and applications.* Berlin: Springer-Verlag, 2012. 210-224 .

Beevolve. *An Exhaustive Study of Twitter Users Across the World.* 2012. http://www.beevolve.com/twitter-statistics.

Berners-Lee, T. 2006. http://www.w3.org/DesignIssues/LinkedData.html.

Berners-Lee, T. et al. "The Semantic Web." *Scientific American* 284 (2001): 34-43.

Blei, D. M, et al . "Latent dirichlet allocation." *Journal of Machine Learning Research*, 3 2003: 993–1022.

Bollen, J. et al. "Twitter mood predicts the stock market." *Journal of Computational Science* 2, no. 1 (2011): 1-8.

Cafarella, M., et al. "Structured Data on the Web." *Communications of the ACM* 54, no. 2 (2011): 72-79.

calculator, Climate. 2014. https://www.gov.uk/2050-pathways-analysis.

Cambria, E. "An introduction to concept-level sentiment analysis." *Advances in Soft Computing and Its Applications.* Springer, 2013. 478–483.

"Citizen Participation." 2012. http://www.businessofgovernment.org/blog/business-government/next-four-years-citizen-participation.

CKAN. 2014. http://thedatahub.org/.

Collier, A., et al. "Understanding and influencing behaviours: a review of social research, economics and policy making in Defra." February 2010. http://archive.defra.gov.uk/evidence/series/documents/understand-influence-behaviour-discuss.pdf.

Connell, N. "Evaluating soft OR: Some reflections on an apparently unsuccessful implementation using a soft systems methodology (SSM) based approach." *Journal of the Operational Research Society*, 2001: 150–160.

Danielson, M., et al. "Cross-disciplinary research in analytic decision support systems." *Proceedings of the 28th International Conference Information Tech. Interfaces.* IEEE, 2006.

DCAT. 2014. http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary.

de Vries, B. et al. "Understanding Design Through Design Support Tools." *Design Research in the Netherlands 2005.* Eindhoven: Eindhoven University of Technology, 2005. 205-214.

Ding, L., et al. "Twc logd: A portal for linked open government data ecosystems." *Web Semantics: Science, Services and Agents on the World Wide Web.* Elsevier, 2011. 325–333.

Druckenmiller et al. "Agent-Based Collaborative Approach to Graphing Causal Maps for Situation Formulation." *Journal of the Association for Information Systems* 10, no. 3 (2009).

Eden, C. "On evaluating the performance of GSS:Furthering the debate." *European Journal of Operational Research*, 2000: 218–222.

Ell, B., et al. "Labels in the web of data." *The Semantic Web – ISWC 2011.* Springer Berlin/Heidelberg, 2011. 162–176.

Fasth, T., Larsson, A. "Portfolio Decision Analysis in Vague Domains." *Proceedings of IEEE IEEM.* IEEE, 2012. 61-65.

Fernandez, M. et al. "Using Social Media To Inform Policy Making: To whom are we listening? ." In *European Conference on Social Media: ECSM*, 174-182. ePub , 2012.

Franco L. A., Montibeller G. "Problem structuring for multicriteria decision analysis interventions." In *Wiley encyclopedia of operations research and management science*, by Cochran et al. (ed). USA: Wiley, 2011.

Franklin, M., et al. “From databases to dataspaces: a new abstraction for information management.” *SIGMOD Record* 34, no. 4 (2005): 27-33.

Friend, J. “The strategic choice approach.” In *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity,Uncertainty and Conflict*, by J., Mingers, J. (Eds.) Rosenhead, 115–150. Chichester: Wiley, 2001.

Fung, A., Shkabatur, J. *iral Engagement: Fast, Cheap, and Broad, but Good for Democracy?* 2012. http://www.archonfung.net/docs/articles/2012/ViralEngagement5.pdf.

Gass, S. I. “"George B. Dantzig". Profiles in Operations Research.” *International Series in Operations Research & Management Science*, 2011: 217–240.

Global. *Global.* 2014. http://www.globalcalculator.org/.

Gonzalez, H. et al. “Google fusion tables: data management, integration and collaboration in the cloud.” *Proceedings of the 1st ACM symposium on Cloud computing.* ACM, 2010. 175–180.

Griffiths, T.L, Steyvers, M. “Finding scientific topics.” *PNAS, 101(suppl. 1)*, 2004: 5228–5235.

Haase, P., et al. “The Information Workbench as a Self-Service Platform for Linked Data Applications.” *Workshop on Consuming Linked Data (COLD 2011).* Bonn: ISWC 2011, 2011.

Halevy, A. Y. “Towards an ecosystem of structured data on the web.” *Proceedings of the 15th International Conference on Extending Database Technology.* EDBT ’12, 2012. 1-2.

Hansard. “Audit of Political Engagement.” 2014. http://www.hansardsociety.org.uk/wp-content/uploads/2014/04/Audit-of-Political-Engagement-11-2014.pdf.

Harnden, R. “The languaging of models: The understanding and communication of models with particular reference to Stafford Beer's cybernetic model of organization structure.” *Systems Practice* 3, no. 3 (1990): 289–302.

Harris, R. *Introduction to Decision Making.* VirtualSalt, 2012.

Hartig, O., et al . “Executing SPARQL Queries over the Web of Linked Data.” *International Semantic Web Conference.* 2009. 293-309.

Heath, T. and Bizer, C. *Linked Data: Evolving the Web into a Global Data Space.* Morgan & Claypool, 2011.

Hedeler, C., et al. “Dataspaces.” In *Search Computing*, by S., Brambilla, M. Ceri, 114-134. Springer-Verlag, 2010.

Hulpus, I., et al. “Unsupervised graph-based topic labelling using dbpedia.” *In Proceedings of the sixth ACM international conference on Web search and data mining, WSDM ’13.* New York: ACM, 2013. 465–474.

Hussain, M., and Howard, P. “The role of digital media.” *Journal of democracy* 22, no. 3 (2011): 35–48.

iResearch. 2014. http://iresearch-reporter.com/.

Isele, R., et al. “LDSpider: An open-source crawling framework for the Web of Linked Data.” *Proceedings of 9th International Semantic Web Conference (ISWC 2010).* 2011.

Jiwani, G.N. *Uncovering the Unknown of Government Policy Decision-making Process at Senior Levels: Multiple Case Study.* University of Washington, 2010.

John, Tony. *What is Semantic Search?* 2012. http://www.techulator.com/resources/5933-What-Semantic-Search.aspx.

Keeney R.L, Raiffa H. *Decisions with multiple objectives: preferences and value tradeoffs.* Cambridge University Press: Cambridge, 1993.

Kouloumpis, E. et al. “Twitter sentiment analysis: The good the bad and the omg!” *Proceedings of the ICWSM.* Barcelona, 2011.

Kullback-Leiber. 2014. http://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence.

Langegger, A. et al. “A semantic web middleware for virtual data integration on the web.” *Proceedings of the 5th European semantic web conference on The semantic web: research and applications.* Springer-Verlag, 2008. 493-507.

Lau, J.H., et al . “Automatic labelling of topic models.” *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1.* Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 1536–1545.

LDA. 2014. http://en.wikipedia.org/wiki/Latent_Dirichlet_allocation.

Leavy, J. *Social Media and Public Policy, What is the evidence? .* Alliance for Useful Evidence report, 2013.

Lenzerini, M. “Data integration: A theoretical perspective.” *21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.* Madison: ACM, 2002. 233-246.

Li, J. et al. “RiMOM: A Dynamic Multistrategy Ontology Alignment Framework.” *IEEE Transactions on Knowledge and Data Engineering.* IEEE Computer Society, 2009. 1218-1232.

Liesiö, J., et al. “Preference Programming for Robust Portfolio Modeling.” *European Journal of Operational Research* 181, no. 3 (2000): 1488-1505.

LinkedData. 2014. http://linkeddata.org/.

Liu, B. “Sentiment analysis and subjectivity.” In *Handbook of Natural Language Processing*, by N. Indurkhya and F. J. Damerau (eds). 2010.

Lloret, E., Palomar, M. “Challenging Issues of Automatic Summarization: Relevance Detection and Quality-based Evaluation.” *Informatica* 34 (2010): 29–35.

LODifier. 2014. http://www.aifb.kit.edu/web/LODifier/en.

Lourenço, J., et al. “PROBE – A Multicriteria Decision Support System for Portfolio Robustness Evaluation.” *Decision Support Systems* 54, no. 1 (2012): 534-550.

Luna-Reyes et al. “Collecting and analyzing qualitative data for system dynamics: methods and models.” *System Dynamics Review* 19, no. 4 (2003): 271-296.

Madden, M. “Older Adults and Social Media, Pew Internet.” 2014. http://pewinternet.org/Reports/2010/Older-Adults-and-Social-Media.aspx.

Magatti,D., et al. “Automatic labeling of topics.” *In Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications, ISDA ’09.* Washington: IEEE Computer Society, 2009. 1227–1232.

Makela, E. *Survey of Semantic Search Research.* 2005. http://www.seco.tkk.fi/publications/2005/makela-semantic-search-2005.pdf.

Mani, I. *Summarization Evaluation: An Overview.* 2014. http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/sum-mani.pdf.

McCrae, J., Spohr,D., Cimiano, P. “Linking lexical resources and ontologies on the semantic web with lemon.” In *The Semantic Web: Research and Applications*, 245–259. Heidelberg: Springer, 2011.

Mei, Q., et al. “Automatic labeling of multinomial topic models.” *In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.* New York: ACM, 2007. 490–499.

Microformats. 2014. http://microformats.org.

Mingers, J., Rosenhead,J. “Problem structuring methods in action.” *European Journal of Operational Research*, 2004: 530–554.

Mintzberg, H. *The Rise and Fall of Strategic Planning.* New York: Prentice-Hall, 1994.

Mitchell, B. *Policy-making process.* 2009. http://www.waterencyclopedia.com/Oc-Po/Policy-Making-Process.html.

Myerson, R.B. *Game Theory: Analysis of Conflict.* Harvard University Press, 1991.

Nenkova, A., Kathleen, K. “Automatic Summarization.” *Information Retrieval* 5, no. 2-3 (2011): 103–233.

Newsblaster. 2014. http://www1.cs.columbia.edu/~sable/research/hlt-blaster.pdf.

Ngomo, A. et al. “LIMES A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data.” *IJCAI.* 2011. 2312-2317.

OR. *What is Operations Research?* 2014. https://www.informs.org/About-INFORMS/What-is-Operations-Research.

Pak, A., Paroubek, P. “Twitter as a corpus for sentiment analysis and opinion mining.” *Proceedings of LREC 2010.* Valletta, 2010.

Phillips, L. “People-centered group decision support.” In *Knowledge-Based Management Support Systems*, by Doukidis et al, 208–224. Chichester: Ellis-Horwood, 1989.

“Politics.” 2014. http://bsa-30.natcen.ac.uk/read-the-report/politics/introduction.aspx.

Puron-Cid, G., et al. “IT-Enabled Policy Analysis: New Technologies, Sophisticated Analysis and Open Data for Better Government Decisions.” *Proceedings of the 13th Annual International Conference on Digital Government Research.* 2012. 97-106.

Radev, D., et al. “Newsinessence: Summarizing Online News Topics.” *Communications of the ACM - The digital society*, 2005: 95-98.

Researcher, NewsFeed. 2014. http://newsfeedresearcher.com/.

ROBUST. 2014. http://www.robust-project.eu/.

Rosenhead. “What is the problem. An introduction to problem structuring methods.” *Interfaces* 26, no. 6 (1996): 117–131.

Rosenhead, J. *Rational Analysis for a Problematic.* Chichester: Wiley, 1989.

Saif, H. et al. “Semantic sentiment analysis of twitter.” *11th Int. Semantic Web Conf.* Boston: ISWC, 2012.

Salo, A., et al. “Improving Resource Allocation with Portfolio Decision Analysis.” *Analytics*, Nov/Dec 2011: 23-26.

Schulte, A., et al. “LDIF - Linked Data Integration Framework.” *Workshop on Consuming Linked Data (COLD 2011).* Bonn: ISWC 2011, 2011.

Semantria. 2014. http://www.sementria.com.

sentistrength. 2014. http://sentistrength.wlv.ac.uk/.

sentiwordnet. 2014. http://sentiwordnet.isti.cnr.it/.

Sterman, J. *Systems Thinking and Modeling for a Complex World.* Boston: Irwin/McGraw-Hill, 2000.

Suchanek F. M., Sozio, M., Weikum, G. “Sofie: a self-organizing framework for information extraction.” *Proceedings of the 18th international conference on World wide web*, 2009: 631–640.

Support. 2014. http://en.wikipedia.org/wiki/Support_vector_machine.

Texalytics. 2014. http://textalytics.com/core/topics-info.

tools. 2014. http://www.ifm.eng.cam.ac.uk/research/dstools/#2b.

“Traffic.” 2014. http://www.comscore.com/.

Tumasjan, A., et al. “Predicting elections with twitter: What 140 characters reveal about political sentiment.” *In Proceedings of the fourth international aaai conference on weblogs and social media.* 2010. 178-185.

Tummarello, G., et al. “Sig. ma: Live views on the Web of Data.” *Web Semantics: Science, Services and Agents on the World Wide Web* 8, no. 4 (2010): 355-364.

van Ossenbruggen, J. et al. “Interactive Vocabulary Alignment.” *International Conference on Theory and Practice of Digital Libraries.* 2011. 296-307.

VoId. 2014. http://vocab.deri.ie/void/guide.

Volz, J. et al. “Discovering and Maintaining Links on the Web of Data.” *International Semantic Web Conference.* 2009. 650-665.

WeGov. 2014. http://www.wegov-project.eu/.

Weimer, D., and A. Vining. *Policy Analysis: Concepts and Practice .* Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

Wheeler, B. *Why not let social media run the country?* 2012. http://www.bbc.co.uk/news/uk-politics-19555756.

Wills, G. J. “Visualizing hierarchical data.” *Encyclopedia of Database Systems.* Springer US, 2009. 3425–3432.

Wilson, T. et al. “Recognizing contextual polarity in phrase-level sentiment analysis.” *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.* Vancouver, 2005.

WordNet. 2014. http://wordnet.princeton.edu/.

YahooAPI. 2014. http://developer.yahoo.com/search/content/V1/termExtraction.html.

Yardi, S., Boyd, D. “Dynamic debates: an analysis of group polarization over time on Twitter.” *Bulletin of Science, Technology & Society* 30, no. 5 (2010): 316-327.