DG DIGIT
ISA^2 Programme

D02.01 Social media analysis guidelines

ISA^2 - Action 2016.04: Participatory Knowledge for supporting decision making

Specific Contract #404

Date: 28/08/2017
Doc. Version: 2.0

Document Control Information

Settings	Value
Document Title:	D02.01: Social media analysis guidelines
Project Title:	ISA2 - Action 2016.04: Participatory Knowledge for supporting decision making
Document Authors:	Massimo Pedroli (KPMG Italy), Ambra Antignani (KPMG Italy), Lorenzo Carbone (KPMG Italy)
EC Project Officer:	Marco Fichera
Doc. Version:	FINAL

Document history:

Revision	Date	Created by	Short Description of Changes
DRAFT	31/07/2017	Massimo Pedroli (KPMG Italy), Ambra Antignani (KPMG Italy), Lorenzo Carbone (KPMG Italy), Antonio Feraco (External Expert)	First release
FINAL	29/08/2017	Massimo Pedroli (KPMG Italy), Ambra Antignani (KPMG Italy), Lorenzo Carbone (KPMG Italy), Antonio Feraco (External Expert)	Final release

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1. EXECUTIVE SUMMARY

The continuous growth of digital Social Media as blogs, Twitter, and Facebook has greatly expanded the reach and speed of propagation and diffusion of information and the adoption of innovations across citizens and business. Primarily, it has done so by increasing the ability of one individual to make his/her thoughts and opinions known to a more extensive community, often beyond their immediate network and community. This same explosion of online activity, enabled by these social web applications, and new computational techniques for the analysis of social network data, has led to the growth of personalized recommendations supporting decision-making processes. In this context is to be noted that information diffusion occurs in different ways, in different formats and from a variety of digital social media (e.g., an unstructured web forum, the structured follower lists on Twitter, or a stream of posts on a blog channel). Moreover, information about different topics diffuses in different ways. For instance, reports on technology diffuse differently than news about popular culture, and opinion, in general, diffuses differently from physical products; additionally, diffusion dynamics may differ both by platform, as well as by topic area, within the platform.

The main barriers for the diffusion of the use of Social Media Analysis across European PAs are summarised below:

- **Lack of an overall vision:** many public administrations are lacking an overall strategy and vision on how to use Social Media (and related information analysis) not only for external or internal communication, but also for decision and policy-making purposes. This is due to several factors including cultural barriers;
- **Lack of a dedicated budget:** the implementation of a social media analysis initiative (including communication campaigns, IT platform, resources etc) and its continuous maintenance and operation is sometimes too expensive for a Public Administration which generally has different priorities;
- **Data privacy and reliability:** identifying an authoritative individual (or a community) as the source of influence or recommendation for PAs decision making processes is critical: information and data collected from social media are not always completely accessible (due to data privacy restrictions) trustable and reliable, also because cannot easily associated to their sources;

- **Technological gaps:** each initiative shall be supported by the right technological platform and/or IT tools. Additionally there is a need of continuous update and maintenance of the APIs exposed by the Social Media which often changes configuration parameters and rights of access.

In this complex context and in order to overcome those barriers, the European Commission has mandated KPMG in the context of the **ISA^2 Action 2016-04 – Participatory Knowledge for supporting Decision-Making** to provide a set of guidelines directed to Public Administrations that are willing to use those huge set and variety of information available via Social Medias. Those guidelines leverage on the outcome of the analysis of existing initiatives of Social Media analysis across EU.

2. INTRODUCTION

2.1. CONTEXT AND SCOPE

This study on "Social Media Analysis Guidelines" is an integral part of the **ISA^2 Action 2016-04 – Participatory Knowledge for supporting Decision-Making** and represents a follow-up of its Task 01 "**Leveraging Social Media full potential to increase citizen engagement and participation in public administrations' decision-making processes"**.

Task 01's intensive data collection activity of Member States' Administrations' best practices embodies the foundation of this present study as these use cases represent the current situation of existing projects/initiatives and the current trends on the use of social media analysis tools by European Member States' public authorities in their decision- and policy-making process.

The **scope of Task 02** is to highlight these use cases' best practices in order to provide a list of guidelines to support the implementation of Social Media Analysis initiatives by EU Public Administrations; in other words, Task 02 **identifies the business and technology requirements and good practices in the use of social media analysis in EU Public Administrations (PAs)**.

The present study's list of social media analysis guidelines will be based on the deeper study of these initiatives in terms of identification of business and technology requirements and good practices:

- **NOMAD** – Policy formulation and validation through non moderated crowd-sourcing. (Coordinates by the University of Aegean);

- **NVWA Initiatives** (Netherlands Food and Consumer Product Safety Authority) - Two initiatives on hold; one initiative implemented "Think before you ink". (Coordinated by NVWA);
- **Step4Youth** - Societal and political engagement of young people in environmental issues. (Coordinated by DRAXIS Environmental S.A.);
- **SENSEI** – Making Sense of Human-Human Conversation Data. (Coordinated by the University of Trento);
- **WeGov** – Where e-Government meets the eSociety. (Coordinated by the University of Southampton);
- **CEF Digital Stakeholder Management Office initiative** – (Coordinated by the EC DIGIT D3);
- **Health Canada Initiatives** – 2 Pilot projects on health topics. (Coordinated by Health Canada);
- **EC DG COMM Social Media initiative** – (Coordinated by the EC COMM A1 Social Media team).

Then, the present study also provides a gap analysis between requirements and the aforementioned good-practices, which will allow future project initiators to benefit from a list of potential actions or projects (*e.g.* development of re-usable tools) that can be carried out to fulfil PAs' requirements.
This Task's analysis will be then based on the three following activities:

- **Identification of business and technology requirements for social media analysis in EU public administrations**;
- **Identification of good practices in this domain;**
- **Gap analysis between use-cases and current practices from the public sector**.

As regards the activities above, during the execution of the study, challenges have emerged in scouting social media initiatives across European Public Administrations (PAs), mainly due to the actual overall limited number of initiatives in this area and to the preference of PAs to use social media for **e-Participation** or **dissemination initiatives**. Moreover, it is worth to mention that identified initiatives have been mainly funded through **EU funding programmes (FP7, H2020)** generally coordinated by multinational Consortia including private companies, research institutes or Academia and then PAs. Therefore, since very few could be identified as "current practices form the public sector", business and technology requirements identified, together with good practices

elements, can be considered as the starting point for all PAs that are willing to use Social Media analytics to support their policy- and decision-making processes.

2.2. STRUCTURE OF THE DOCUMENT

This report represents the final deliverable of Task 02. The report contains three main sections, structured according to the approach to this second part of the study, each section detailing the main findings and forming the basis for the next steps and sections:

- **Methodological Approach** – presenting the methodological approach used in pursuing the study. The methodology identifies best practices elements and critical success factors for social media analysis initiatives from current trends and use;
- **Guidelines for the Implementation of Social Media Analysis Initiatives** – providing practical guidelines for future implementation of social media analysis initiatives. The content of this section has been developed along five main steps: 1) How to define the overall vision/strategy of the initiatives; 2) Design the implementation processes; 3) Driver for the data analysis; 4) Technology, tools and capabilities; 5) Initiatives' KPIs and KSFs;
- **Conclusion and next steps** – describing main results/evidences of the study and providing insights on possible future developments.

3. METHODOLOGICAL APPROACH

This section provides a description of the approach applied in order to ensure completeness and full coverage of the related task, so as to guarantee the achievement of the objectives of the study, and namely:

- the identification of a list of good practices and guidelines on the use of Social Media for EU PAs, and related supporting IT tools;
- the definition of business and technology requirements for social media analysis in EU Pas.

3.1. IDENTIFICATION OF BEST PRACTICES ELEMENTS AND CRITICAL SUCCESS FACTORS FROM USE CASES

One of the main goals of this study refers to the identification of good practices and critical success factors that could support PAs in the implementation of social media initiatives.

In this regard, the output of the primary and secondary data collection activities carried out during Task 01 (e.g. interviews, analysis of previous studies, etc.) and detailed described within **"D01.01 Social media use-cases, current usage and trends from public administrations"** has been extensively leveraged in order to identify, among selected initiatives, re-usable technologies and replicable practices that could allow the spread of Social Media Analytics as a means to support a more participatory type of government and to enhance a democratic participation of stakeholders in policy- and decision-making.

Amongst activities carried out within Task 01, the following set the ground for the definition of a shortlist of initiatives whose features allowed identifying good practices and key success factors that could be replicated:

- **Data Collection** – primary and secondary data collection activities were executed to collect relevant information for the study. In detail, a long list of citizen-sourcing initiatives was identified;
- **Selection and Assessment Model definition** – a clear set of criteria (both qualitative and quantitative) was defined to sort initiatives relevant for the study. In addition, an assessment model was developed in order to perform a detailed analysis of selected initiatives on the basis of qualitative and quantitative criteria;
- **Identification of initiatives eligible for future pilots** – within this phase, after the shortlisting and detailed analysis of relevant projects, initiatives were classified based on their level of re-use and a priority was assigned in order to identify initiatives eligible for future implementations and piloting by DIGIT ISA Unit.

Within this context, it has to be noted that all analysed initiatives allowed identifying useful elements that contributed to the purposes of this Study. Such elements have been further elaborated for developing the set of Guidelines outlined in the following paragraphs, in order to support and foster the implementation of Social Media initiatives by EU PAs.

Indeed, the in-depth analysis of such initiatives has led to the identification of **several elements that can be considered as good practices**, together with a number of key success factors of those initiatives that can be generalized and taken as reference. Such elements mainly refer to the following areas that can be considered key for the implementation of projects related to Social Media Analysis for supporting policy- and decision-making[1]:

- **Overall vision and strategy**, attaining at how the overall concept and strategy of the initiative, its main goals and purposes and how they have been communicated and disseminated both to the political stakeholders and the wider public. This includes also stakeholders' engagement strategies, when developed;
- **Implementation process**, in terms of definition of clear steps for the effective and successful start-up and execution of the initiative. Implementation processes mainly refer to the identification of phases and/or steps that a PAs should follow when implementing a Social Media Analytics initiative;
- **Data analytics**, referring to the methodology and tools utilized to collect/cleanse/analyse/process information available on Social Media in order to obtain relevant result for policy- and/or decision-making purposes;
- **Technology and tools**, in terms of main features and functionalities of the technology application developed (or purchased) to support the initiative, with special regard to re-usability, openness and modularity of the technology.

Within the following paragraph 4, "Guidelines for the implementation of Social media analysis initiatives", good practices and key success factors related to the main areas outlined above are detailed in order to provide guidelines to be used for the potential implementation of future initiatives and pilots in the Social Media Analysis for EU PAs.

[1] See D01.01, "Social media use-cases, current usage and trends from public administrations", p. 57

3.2. EMERGING BUSINESS NEEDS FROM USE CASES AND MEMBER STATES' FEEDBACK

The initiatives analysed during the first phase of the study have provided, among others, inputs on the business needs that a successful social media analysis initiative should satisfy. Similarly, the member states contacted to collect information on their experiences and requirements in this domain have contributed to such analysis based on their business objectives, the initiatives that they have launched or have planned to launch, and the possible resulting gaps towards the expected results. Such needs have then been taken into account to qualify the guidelines (see Chapter 4, "Guidelines for the implementation of Social media analysis initiatives").

Business needs have been classified in the same four areas in which the good practices were grouped, as mentioned above:

- Overall vision and strategy;
- Implementation process;
- Data analytics;
- Technology and tools.

The emerging business needs have then been defined through the following actions:

1. **Identifying the possible gaps towards the best practices** in the other initiatives analysed during the first phase of the project;
2. **Elaborating on the challenges and needs** explicitly mentioned by the stakeholders of the analysed initiatives;
3. **Elaborating on the business needs** mentioned by the member states contacted in this phase.

For each of those areas, a second-level, more detailed clusterisation of the needs has also been provided, when possible. Table 1 below lists the business needs collected through the process mentioned above, each described with the following attributes:

- **ID:** an identifier of the business need obtained through the analysis of the Big Data pilots. This ID will link each business need with the resulting guidelines described in Chapter 4;
- **Area:** a classification of each business need in the four areas mentioned above, with the indication of sub-areas where necessary;
- **Reporter:** the initiative or member state representative reporting the business need;

- **Need description:** synthetic description of the business need.

Table 1 – Emerging business needs

ID	Topic	Reporter	Need description
1	Vision and strategy / Engagement	NOMAD	Overcome cultural barriers in the public administrations on the use and potential of social media (and related data analysis). Improve public administrations resources' engagement.
2		all	Need to raise the awareness of PAs on the use of such initiatives for policy-making purposes, exploiting "passive" crowdsourcing along with "active" crowdsourcing
3		Health Canada, WeGov	Set up efficient governance mechanisms in order to ensure the proper follow-up and implementation of the initiative, with possible iterative developments upon ongoing interactions with public stakeholders.
4	Implementation / process	all	Define a well-structured set of processes and procedures for the management of the initiative, to optimise time and use of resources and effectively satisfy its analytical needs.
5	Implementation / Data privacy	NOMAD, WeGov	Data privacy regulations limit the access to and use of online personal data and the variation of data protection legislation among EU MSs makes it difficult to draft common privacy policies for cross-country projects.
6		WeGov, SENSEI	Clarify the limitations to the use of data deriving from the Terms and Conditions and policies of the single Social Media platforms.
7	Implementation / Engagement	NOMAD, SENSEI, Health Canada	Increase the co-operation with and the engagement of the media communication companies, in order to obtain their consent to access closed data.
8		SENSEI	Bring together institutions covering the various competencies needed for such initiatives (communication, technology, policy making…), also from different countries for EU-wide projects.
9	Implementation / Funding	All	Adequate funding (i.e. using EU funding programmes) needed to integrate local / national resources, especially for EU-wide projects.
10	Implementation / People	NOMAD	Sufficient time and resources needed on the PAs side to support the initiatives.
11		NOMAD	Time and effort needed to provide day-by-day support to the people/resources engaged, to convey the concept, to make them to be familiar with the tool, and to implement it in a structured way.
12		NOMAD, NVWA, Health Canada	PAs need to involve specialists/experts with adequate skills in data analytics, digital and regulatory affairs (i.e. policy).
13	Data analytics	WeGov	Identification of the appropriate data source(s) based on the data needed and on the purposes of the initiative.
14		SENSEI	A (wider) availability of metadata (in particular, geospatial information) for the clusterisation of data would improve the quality of the analysis especially in terms of opinion trends monitoring.
15		All	Availability of software tools, modules or services with adequate social media analytics capabilities.
16		STEP4YOUTH	Make sure that the technology used detects the reliability of the analysed data (e.g. avoiding trolls).

ID	Topic	Reporter	Need description
17		CEF Stakeholders Engagement, NOMAD	In case of targeted selection of the data providers, make sure that the most relevant ones (e.g. “influencers”) are selected and monitored.
18		NOMAD	Significant computational capacity is needed to support the execution of simultaneous analysis on numerous users and in different languages.
19	Technology and tools	NOMAD	Ensure continuous maintenance of the SW platform that crawls data coming from proprietary external data sources (i.e. LinkedIn, Facebook, etc.) having a quick timeframe in changing their APIs to share data
20		STEP4YOUTH	Develop user-friendly yet effective tools that may cope with the limited familiarity with IT technology in public administrations, especially at local level
21		STEP4YOUTH	The technology and tools supporting the analysis must be open source and reusable.
22		CEF Stakeholders Engagement	Automatise the data collection process even for initiatives targeting a limited number of data sources / monitored accounts.
23		NVWA	Need for tools, either COTS or autonomously developed, customizable in accordance to the evolving requirements of the analysis.
24		NVWA	In case of use of a social media analysis infrastructure, this should satisfy at least the following requirements: - clear SLAs in terms of uptime / performance - adequate information security - customization possibilities - software compatible / interoperable with the IT architecture of the PA.

In further synthesis, we could say that the needs expressed by the stakeholders of the analysed initiatives and the representatives of the member states are focused around the following areas:

- Have a clear framework in terms of **data privacy** regulations and policies;
- Get access to **reliable data**, also **closed** ones, and **metadata** allowing proper clusterisation according to the analytical needs;
- Raise the **awareness and engagement of public administrations** potentially interested in social media analysis initiatives;
- **Improve the digital skills** of people to be involved in such initiatives;
- Exploit **open social media analysis IT platforms,** providing **adequate** and **customisable data analytics features** and ensuring a **reliable service level**.

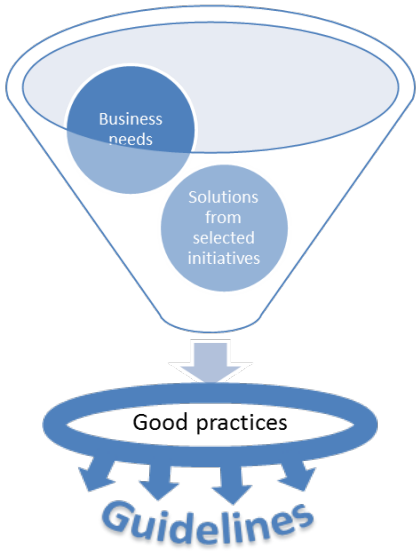
4. GUIDELINES FOR THE IMPLEMENTATION OF SOCIAL MEDIA ANALYSIS INITIATIVES

4.1. INTRODUCTION

The objective of this step of the study is to develop guidelines for social media analysis initiatives based on good practices elements implemented in the context of the initiatives analysed in the previous phase of the study.

In particular, this step of the analysis maps the business needs identified in Paragraph 3.2 against the shortlisted initiatives and highlights how the initiatives have addressed such needs. This has allowed representing how the specific business challenges that a Social Media initiative shall satisfy can be overcome. From here, through an inductive process, the study defines the guidelines for future similar initiatives.

Figure 1 - Methodological approach for the identification of guidelines



While the method applied has followed the inductive approach mentioned above, this chapter renders the output of the analysis organised by guidelines, each presented in a dedicated sheet. This format is chosen so to provide immediately applicable guidance, to the design of future social media analysis initiatives, allowing stakeholders to identify for each identified business challenge a reference to how any existing initiative had already implemented that specific guideline.

Each guideline is provided through a sheet identifying the related business need(s) that the implementation of the guideline contributes to satisfy, and a section describing which initiatives

have best put in place such guideline, and through which good practice element. Additionally, other references coming from scientific literature are also mentioned, where applicable.

5. HOW TO DEFINE THE OVERALL VISION/STRATEGY OF THE INITIATIVE

<table>
<tr><th colspan="4">GUIDELINE 1</th></tr>
<tr><td>Area</td><td colspan="3">Vision and strategy</td></tr>
<tr><td>Topic</td><td colspan="3">Engagement</td></tr>
<tr><td>Related business need</td><td colspan="3">1. Overcome cultural barriers in the public administrations on the use and potential of social media (and related data analysis). Improve public administrations resources' engagement.</td></tr>
<tr><td>Description</td><td colspan="3"><u>Execution of Pilot Projects involving EU Public Administrations</u>
In order to gradually make Public Administrations' professionals more confident about how to use Social Medias and their data, the execution of a set of local Pilots in domain of interests could be foreseen. This action could improve to increase Social Media related skills inside the PA and could help to overcome existing cultural barriers.</td></tr>
<tr><td rowspan="3">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented</td></tr>
<tr><td>STEP4YOUTH</td><td>  </td><td>The project has run a significant range of different pilots in 4 countries:
<u>The Association of the Communities of Locride area</u>: It consists of 42 municipalities located in the Locride area, Italy, sharing the goal to promote development and democracy;
<u>Mollet del Vallès</u>: A city of the metropolitan area of Barcelona that has won the European Green Leaf Award;
<u>Valdemoro city</u>: A municipal district, located in the Southern zone of the autonomous community of Madrid, Spain;
<u>The Region of Crete</u>: A second grade local self-government authority in Southern Greece
<u>Hatay</u>: One of the largest cities located in Southern Turkey, on the Mediterranean coast, that has recently become a Metropolitan Municipality
Such extended piloting approach was also meant to foster:
the use of a cloud eParticipation SaaS platform used by young citizens in selected pilot countries;
societal and political participation of young people in the decision-making process on environmental issues;
a sustainable network of users (young citizens, decision makers, stakeholders).</td></tr>
<tr><td>NOMAD</td><td>  </td><td>The project has run 3 pilots:
- Greece (Hellenic Parliament), on Energy Planning;
- UK (Critical Publics, a UK based intelligence consultancy), on immunotherapy;
- Austria (Austrian Parliament), on open data.</td></tr>
</table>

			Common aim of these pilots was, among others, to run an awareness campaign targeted to specific points of misunderstanding/mistaken information. To this end, pilot follow-up and dissemination activities were executed.
Other	N.A.		

GUIDELINE 2			
Area	Overall vision and Strategy		
Topic	Engagement		
Related business need	2. Need to raise the awareness of PAs on the use of such initiatives for policy-making purposes, exploiting "passive" crowdsourcing along with "active" crowdsourcing.		
Description	***Set up a dissemination strategy including dissemination materials and actions*** One of the main elements that can support the implementation of Social Media Analysis initiatives is the creation of awareness of PAs and other target groups towards the potentialities of such initiatives. The design of a dissemination strategy is the main tool that can support the achievement of such objectives and should include: • the deployment of a Dissemination plan; • the use of appropriate formats and effective communication procedures; • the equal and timely distribution of information on the project among all partners and stakeholders; • the minimization of redundancies and the maximization of accessibility and ease of use of information transmitted. The Dissemination plan should foresee the production of dissemination materials and the deployment of dissemination actions, such as: • Dissemination materials - Definition of a Visual identity of the initiative (initiative's logo); - Creation of a dedicated website that guarantees equal access to all stakeholders and clearly explains background and objectives of the project together with the possibility of downloading relevant material; - Distribution of promotional material (videos, newsletters, press releases, Leaflets and posters). • Dissemination activities - Creation of a Network of stakeholders helping in the dissemination of all activities and results achieved during the initiative among all relevant stakeholders (PAs and target citizens groups); - Mass media communication in order to inform the general public about the initiative and its objectives, thus raising awareness and building consensus at a broader level; - Press releases distributed in English and, when possible, in all languages of partners involved in the projects, in order to ensure the maximum degree of accessibility; - Publications: to the extent to which Academia and Research institutes are concerned, the preparation of publications in scientific and peer-review journals would raise awareness among the scientific community thus helping to enhance reliability of the initiative.		
Reference initiatives	**Name**	**Nationality**	**How was the practice implemented?**
	STEP4YOUTH		A full, integrated dissemination plan was developed, covering aspects such as: - the visual identity of the project - the proper design of the website - the use of social media as dissemination channel - multimedia promotional materials (videos, newsletters, press releases, posters and leaflets)

			- the presence on multiple channels (mass media, network of interest, scientific publications, events) and including the evaluation of its effectiveness.
	NOMAD		Dissemination activities carried out during the project included a number of papers published on journals and conference proceedings. Such documents were then made available on NOMAD website.
Other	N.A.		

GUIDELINE 3			
Area	Overall vision and Strategy		
Topic	Engagement		
Related business need	3. Set up efficient governance mechanisms in order to ensure the proper follow-up and implementation of the initiative, with possible iterative developments upon ongoing interactions with public stakeholders.		
Description	***Design and implement a Governance model*** In order to set up a strategy that effectively supports the successful implementation of the initiatives, the following principles should be taken into consideration: **1. Identify a Social Media team:** identify a dedicated team inside the organisation, responsible to coordinate the implementation and monitoring of all the activities related to the Social Media Initiative to be carried out by the PA; **2. Focus on Issues of Interest**: prior to launch a social media analysis initiative, the issue(s) on which the PA is willing to focus should be clearly identified and should refer to topics of interest both for the PAs and for targeted citizens. It is important to identify concrete issues in order to avoid disengagement and a low amount of data available. Citizens should feel close to the problem discussed in order to enhance commitment and engagement. In this regard, PAs that are willing to implement a social media analysis initiative should undertake some research to ensure that the focus is on issues concerning the targeted groups, to foster greater interest and engagement; **3. Provide clear purposes**: clearly define guidelines on how participants' input will be used and on the type of feedback that will be provided: • **Promote Empowerment & Leadership**: involve users from targeted group in the definition of the process in order to build trust and show how actions are actually taken on the basis of provided contributions and opinions; • **Provide feedback**: in order to build trust around the initiatives and enhance their effectiveness, it is important that the PAs determine ways for communicating to the large public how the participation made a difference and how the information has been used. In particular, it is important to provide feedback on what actions have been taken upon contributions and opinion gathered; **4. Establishing partnership**: when launching a Social Media Analysis initiative, the establishment of partnership with relevant organisations for the topics of interest (e.g. NGOs) can help to promote inclusive and participatory environment avoiding the creation of wrong expectations. Moreover, usually local organisations have strong ties with the local communities and can support the creation of positive perceptions around the initiatives.		
Reference initiatives	**Name**	**Nationality**	**How was the practice implemented?**
	STEP4YOUTH		The project focused specifically on environmental issues; it may then count on the cooperation of young citizens specifically engaged on a policy topic enjoying *per se* wide public awareness. Special care was also put on the bi-directional supply of information, informing the participants to the

			project on what other people are saying on the specific issues of interest, filtering information from noisy content in social media and web streams, and providing it translated in their own language. Furthermore, the project gave the participants the opportunity to bring specific issues to the attention of policy makers.
	NOMAD		For each Pilot implemented, the following actions were implemented at a governance level: - Involvement of policy makers for the selection of the policy topics to investigate in order to focus the social media analysis on topics of interest of the owners; - Organisation of focus group with stakeholders in order to evaluate the effectiveness of the initiative on the basis of a multi-perspective framework (including political, crowd-sourcing and diffusion elements); - Use of both qualitative and quantitative techniques in order to evaluate stakeholders' feedbacks.
	WeGov		Scientific findings of the concept, technical features, functionality and impact of the WeGov solution have been published in various conference proceedings and/or scientific journals.
Other	N.A.		

5.1. DESIGN THE IMPLEMENTATION PROCESSES

GUIDELINE 4	
Area	Implementation
Topic	Implementation process
Related business need	4. Define a well-structured set of processes and procedures for the management of initiative, to optimise time and use of resources and effectively satisfy its analytical needs.
Description	***Design supporting processes and procedures*** When considering the concrete implementation of an initiative, it is key to clearly define a business process to support the project activities in day-by-day operations. Once that the process is clearly structured it is possible to streamline the related flow of activities. ***A. Overall process*** A potential methodological framework to set up the implementation process in social Media Analytics should consider the following 4 steps: 1. **Monitoring people needs, opinions and proposals**: to this end, a crawling tool should be implemented, searching for information from different data sources such as micro-blogging sites (for example, Twitter) or blogs (blogger, WordPress, Typepad, LiveJournal, etc.), video sites; 2. **Analysis of information**: this includes all analysis of information hidden in conversation already present on Social Media in order to obtain relevant results for policy purposes; 3. **Reporting:** elaboration and representation of data collected in order to provide relevant results**.** This could be done, for example, by the use of visual analytics so as to present useful insight gained from the analysis phase; 4. **Implementation of actions**: once that information are analysed and results are obtained, actions should be implemented within the policy agenda; 5. **Communication of results**: to spread the results of the initiatives and increase engagement of citizens. ***B. Data Analysis process*** Under this general process framework the data analysis process of a Social Media Analysis initiative could be effectively implemented in the following steps: 1. **Set-up of the "domain model"** in order to identify the policy objectives and areas of interest of the involved policy makers. The "domain model" consists in a representation of the main domains that the policy-makers intend to address through a policy (e.g. energy domain, education domain, health domain), as well as relations among them, in a tree structure; this could be done using a graphical modelling tool; 2. **Definition of the "policy model"** based on the above domain model. The policy model represents the public policy to be investigated in the social media, always in a tree structure, with a number of "policy statements" associated with one or more nodes of the policy model and for each of them a set of positive or negative "arguments". This could be done using a graphical modelling tool; 3. **Identification of social media sources** (e.g. political blogs, websites, Twitter, Facebook, etc.) to be crawled in order to collect relevant content about the domain or public policy in scope. The identification of sources should be carried out by the policy-makers; 4. **Sources crawling** with the use of an ICT tool in order to find relevant content against the domain or the public policy of interest and **fine-tuning of requirements** of the model (e.g. automated generation of keywords to enrich the policy model). Processing of collected content using opinion mining techniques and presentation of results in a reporting form. Both for the overall process and the data analysis process, the introduction of a stage-gate process where each step is undertaken only after having positively assessed outputs and results of the previous ones, could increase the effectiveness and reliability of results of the initiative.

	Name	Nationality	How was the practice implemented?
Reference initiatives	NOMAD		Design and implementation of an overall approach based on 4 phases (Listen, Analyze, Receive, Act) and of a methodology for data collection structured in 4 steps as well (building the domain model, building the policy model, identification of social media sources, source crawling and data processing)
	Health Canada		The key success factor of the project is the implementation process in terms of setting clear steps and deliverables. The following 4 phases are foreseen: 1 **"Proof of Technology":** answers to the question "Does the Technology actually works?"; 2 **"Proof of Value":** assesses the value of detection for each use case (characterisation model); 3 **"Prioritisation":** appropriately address priorities after detection; 4 **"Integration":** data integration as a result of social media analysis Each step is assessed against the expected delivery value and goals.
Other	Loukis, E., et al., Promoting Open Innovation in the Public Sector Through Social Media Monitoring, Government Information Quarterly (2016)		

GUIDELINE 5			
Area	Implementation		
Topic	Data privacy		
Related business need	5. Data privacy regulations limit the access to and use of online personal data and the variation of data protection legislation among EU MSs makes it difficult to draft common privacy policies for cross-country projects		
Description	One of the major focus for initiative dealing with personal data (and even more, when the mass of data potentially treated concerns a significant number of data owners) is how to deal with data privacy issues. According to the EU Data Protection Directive (95/46/EC), "personal data" is *"any information relating to an identified or identifiable natural person"*. The core of the issue is therefore linked to the possibility to identify a person through the data managed, in which case the data treatment falls under the provisions of applicable data protection regulations. There are several possible actions to be taken, that may be performed concurrently or alternatively, depending on the nature and scope of the data, in order to make sure that provisions for personal data protection are fulfilled: • Using data that may not be directly traced back to individuals who may be identified through them or with the support of additional information; • Applying data anonymisation techniques; • Implementing disclaimers on the use of personal data; As for data anonymisation, however, the following caveats must be taken in consideration: • Anonymisation may remove relevant information, thus making the data collected unfit for the purposes of the analysis; • Complete anonymisation may not be guaranteed, as it is not always possible to fully determine which data must be removed at this aim.		
Reference initiatives	**Name**	**Nationality**	**How was the practice implemented?**
	SENSEI		The project has paid specific attention to preserve the distribution of token within anonymized concept values.

	WeGov		Data anonymisation measures and techniques have been studied in the project, but such measures and techniques do not fully protect against privacy issues. We Gov has also considered other turnaround solutions which allow identifying and analysing the user's profile, in particular the possibility to rely on the voluntary disclosure of personal information by the engaged citizens.
Other	N.A.		

<table>
<tr><th colspan="4">GUIDELINE 6</th></tr>
<tr><td>Area</td><td colspan="3">Implementation</td></tr>
<tr><td>Topic</td><td colspan="3">Data privacy</td></tr>
<tr><td>Related business need</td><td colspan="3">6. Need to clarify the limits of the use of data deriving from the Terms and Conditions and policies of the single platforms</td></tr>
<tr><td>Description</td><td colspan="3">The various data sources apply diverse data privacy policies, depending on the applicable legislation and on the nature of the information treated. Considering the difficulty of developing an autonomous data privacy policy compliant with single provision of each source, it is recommended to develop a common framework for understanding the kind of usage of personal data allowed by each source, classify the source accordingly and treat the pertinent data based on the standard usage levels defined.
A structured analysis of the approach applied by similar past projects using the same data sources, to be considered as references, should also be taken in consideration.</td></tr>
<tr><td rowspan="2">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>SENSEI</td><td></td><td>A compared analysis of the data policies of the various concerned platforms and online media has been conducted, under the aspects of:
• Requirement of written approval;
• Right to use / edit / redistribute;
• Possibility to exploit User Generated Content (UGC).
Accordingly, standard Data Usage Levels (DULs) have been defined, and each data source has been assigned a DUL.</td></tr>
<tr><td>Other</td><td colspan="3">N.A.</td></tr>
</table>

<table>
<tr><th colspan="4">GUIDELINE 7</th></tr>
<tr><td>Area</td><td colspan="3">Implementation process</td></tr>
<tr><td>Topic</td><td colspan="3">Engagement</td></tr>
<tr><td>Related business need</td><td colspan="3">7. Increase the co-operation with and the engagement of the media communication companies, in order to obtain their consent to access closed data</td></tr>
<tr><td>Description</td><td colspan="3">Rules for accessing and using platforms' APIs are becoming more and more restrictive, also for generally more open ones such as Twitter. For example, Twitter's terms and conditions now state: "If you want to reproduce, modify, create derivative works, distribute, sell, transfer, publicly display, publicly perform, transmit, or otherwise use the Services or Content on the Services, you must use the interfaces and instructions we provide, except as permitted through the Twitter Services, these Terms, or the terms provided on dev.twitter.com."
This is even more necessary for those platforms that do not inherently publicly display the information posted by users, such as Twitter, but apply more restrictive privacy terms, as in the case of Facebook. Facebook API is indeed far more limiting than Twitter's, allowing queries to public groups but not personal profiles.
Ad-hoc agreements with the social media platforms should then be reached to access and treat closed data in particular.</td></tr>
<tr><td rowspan="2">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>NOMAD</td><td>  </td><td>During the project an in-depth study of the publicly available software modules and interfaces of the Web 2.0 applications that constitute data sources was carried out. The purpose of the study was the identification of the capabilities that the Social media platforms offer to retrieve content, types of information they offer and standards to structure and describe content. As a result, it was possible to determine how the NOMAD tool components would have interacted with each platform to extract content relevant for the project.</td></tr>
<tr><td>Other</td><td colspan="3">N.A.</td></tr>
</table>

<table>
<tr><th colspan="4">GUIDELINE 8</th></tr>
<tr><td>Area</td><td colspan="3">Implementation</td></tr>
<tr><td>Topic</td><td colspan="3">Engagement</td></tr>
<tr><td>Related business need</td><td colspan="3">8. Bring together institutions covering the various competencies needed for such initiatives (communication, technology, policy making…), also from different countries for EU-wide projects
12. PAs need specialists with adequate skills in data analytics and regulatory affairs</td></tr>
<tr><td>Description</td><td colspan="3">Social media analysis projects are typically multidisciplinary, as they involve skills related at least to the following areas:
• <u>Political sciences</u>: this specific set of competences should allow researchers using social media analytics identifying and selecting the sets of information relevant for public decision making, based on pre-defined parameters;
• <u>Social media</u> / <u>social networking</u>: the knowledge and understand of social media dynamics is a basic background for such projects: it is necessary to understand what kind of users are more likely to be found and sampled on a platform vs others, how to identify and select influencers and trend-setters, understand the semantics typically used on such platforms to identify common patterns, etc.;
• <u>Data analytics</u>: social media analysis is one of the fields the data collected in social media analysis require treatment at different degrees of complexity, depending on the scope of the sample and the type of data (e.g. qualitative vs. quantitative, structured vs. unstructured, etc.). Due to the potentially significant volume of data treated, in particular, it is a typical use case of big data analytics, where political scientists increasingly need to understand how to access and use such data and the tools treating them. Further details on data analytics specialisations covered in social media analysis projects are provided in <u>section 5.2</u>;
• <u>Data privacy</u>: social media analysis potentially manages personal data, thus involving serious privacy issues; special care must be put in analysing, understanding and dealing with such issues, which implies the need to have onboard legal skills pertinent to this specific area, coordinated with data processing techniques (data anonymisation).
<u>Therefore, when implementing a Social Media Analysis project, a PA should involve resources covering the skills mentioned above.</u></td></tr>
<tr><td rowspan="3">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>WeGov</td><td>  </td><td>The consortium represented in the project WeGov is constituted of seven experts in the eParticipation community (Hansard, Gov2u, GFI), the government consulting (GESIS), and the ICT research and development academic community (University of Southampton – ITInn , OU – Knowledge Media institute, University of Koblenz-Landau – WeST), with an outstanding international reputation and skills in advanced open social networking platforms. The participation of commercial and NGO partners with exploitation interests in both IT and non-IT markets associated with eGovernment provides a strong assurance that the project results will be widely exploited.</td></tr>
<tr><td>SENSEI</td><td>  </td><td>The project has been led through the partnership among various European and worldwide-known universities and companies: Università di Trento (Italy), Université d’Aix-Marseille (France), University of Sheffield and University of Essex and Teleperformance (world leader for contact centres) (UK), and WEBSAYS (SME, Barcelona); the</td></tr>
</table>

			University of Trento being the project coordinator and director of the scientific and technological agenda.
Other	N.A.		

GUIDELINE 9			
Area	Implementation		
Topic	Funding		
Related business need	9. Adequate funding is needed to integrate local / national resources, especially for EU-wide projects 10. Sufficient time and resources needed on the PAs side to support the initiatives		
Description	As social media data becomes increasingly commodified, challenges of free access for non-profit research compound. This may lead to the stratification of data haves and have-nots with academics in the latter category. This is true, even though researchers have tools freely available for social media analysis, as they are limited by platform API agreements. To overcome the potential "Big Data digital divide" between academics and large corporations, it is necessary to acquire the resources needed to overcome the existing barriers in exploiting the potential of big data analysis. The main barriers are obtaining permission of social media platforms to use their data and having the technical capabilities needed to interact with the data collected. In order to cover the cost of developing and / or using the computational resources required for such tasks, academic institutions and public administrations may make use at various degrees of the funding opportunities offered by public institutions. At European level, in particular, the frameworks available at this aim are: • **Horizon 2020**: it is the financial instrument implementing the Innovation Union, a Europe 2020 flagship initiative aimed at securing Europe's global competitiveness through research and innovation. Along with other connected measures it aims at breaking down barriers to create a genuine single market for knowledge, research and innovation. The process to obtain such funding is streamlined and easily accessible; • **Structural funds and cohesion funds**: under the objective of fostering transnational cooperation, for example, such funding opportunities may be exploited to build up international consortia around research projects, including those on social media analysis (in fact, several of the analysed projects were run by such consortia). In combination with ad-hoc funding, projects with complex may consider exploiting capabilities made available for common, shared use for a series of initiatives. This may apply, for example, for big data analysis capabilities.		
Reference initiatives	**Name**	**Nationality**	**How was the practice implemented?**
	NOMAD, SENSEI, STEP4YOUTH, WeGov	Various	The consortia managing the projects have made use of various funding opportunities offered by the EU, in particular: - the Framework Programme 7 (FP7 ICT), which supports exploring and testing new approaches, methods and techniques to extract, interpret and exploit information from unstructured multilingual and/or media sources, yielding actionable knowledge; - Horizon 2020 (EU.3.6. - SOCIETAL CHALLENGES), aiming at fostering inclusive and innovative societies.
Other	Mylynn Felt: Social media and the social sciences: How researchers employ Big Data analytics. Sage Journals, April 29, 2016		

<table>
<tr><th colspan="4">GUIDELINE 10</th></tr>
<tr><td>Area</td><td colspan="3">Implementation process</td></tr>
<tr><td>Topic</td><td colspan="3">People</td></tr>
<tr><td>Related business need</td><td colspan="3">11. Time and effort are needed to provide day-by-day support to the people/resources engaged, to convey the concept, to make them to be familiar with the tool, and to implement it in a structured way.</td></tr>
<tr><td>Description</td><td colspan="3">One of the main issues referred to by the policy makers potentially interested in social media analysis initiatives is the lack of time and resources to support the initiative in the structured way.
In this regard, one of the main challenges to overcome is providing day-by-day support to the people/resources engaged. It takes a lot of effort to convey the concept and to make them to be familiar with the analytical tool used, and to implement it in a structured way.</td></tr>
<tr><td rowspan="2">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>NOMAD</td><td>  </td><td>Dedicate (part of) a project resource specifically to the interactions with and engagement of the public administrations involved in the project.</td></tr>
<tr><td>Other</td><td colspan="3">N.A.</td></tr>
</table>

5.2. DRIVER FOR THE DATA ANALYSIS

In addition to the Data Analysis process described in the previous paragraph, in order to provide the end user with a better understanding on how to define the drivers for data analysis in Social Media, it is worth to list some definition in use among data analysis for social media:

- **Natural language processing (NLP)**: NLP is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computers and human (natural) languages. Specifically, it is the process of a computer extracting meaningful information from natural language input and/or producing natural language output.
- **News analytics**: it is the measurement of the various qualitative and quantitative attributes (sentiment, relevant, novelty, etc.) of textual (unstructured data) news stories.
- **Opinion mining**: it is the area of research that attempts to make automatic systems to determine human opinion from text written in natural language.
- **Scraping / data gathering**: it is the process of collecting online data from social media and other Web sites in the form of unstructured text and also known as site scraping, web harvesting and web data extraction.
- **Sentiment Analysis**: it is the ability to capture the sentiments of the general public about social events, political movements, marketing campaigns, and product preferences.

- **Text analytics**: it involves Information retrieval (IR), lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization and predictive analytics.

It is also worth to list some of the challenges faced by actors in this field, in order then to analyse the specific requirements identified within the scope of this study. These challenges include:

- **Scraping / Data gathering**: although social media data is accessible through APIs, due to the commercial value of data, most of the major sources such as Google, Facebook and Tweeter are making it increasingly difficult to obtain comprehensive access to their raw data. Sources like Thomson Reuters and Bloomberg typically charge for premium access to their data.
- **Data cleansing**: it is the process of cleaning unstructured textual data (i.e. normalising the text), especially high – frequency streamed real-time data.
- **Holistic data sources**: it consists in the merging of different data sources, i.e. bringing and combining data from social media, real-time market and customer data and geo-spatial data for analysis.
- **Data protection**: once the resource is created, the data needs to be secured, ownership and IP issues resolved, and users provided with different level of access, in order to avoid unexpected management of data from unauthorized users.
- **Data analytics**: it is a sophisticated analysis of social media data for opinion mining. It raises many challenges, mainly due to the ontology used as a knowledge base: foreign languages, foreign words, slang, spelling errors and natural evolving of language.
- **Analytics dashboards**: several social media platforms require users to write APIs to access feeds or program analytics models in a programming language.
- **Data visualisation**: visual representation of data is needed in order to provide the end users with a clear and effective graphical mean.

<table>
<tr><th colspan="4">GUIDELINE 11</th></tr>
<tr><td>Area</td><td colspan="3">Data Analytics</td></tr>
<tr><td>Topic</td><td colspan="3">Planning – Targeting the right sources</td></tr>
<tr><td>Related business need</td><td colspan="3">13. Identification of the appropriate data source based on the data needed and on their intended use
17. In case of targeted selection of the data providers, make sure that the most relevant ones (e.g. “influencers”) are selected and monitored.</td></tr>
<tr><td>Description</td><td colspan="3">Once the vision/strategy and the implementation processes have been set, and the data analysis process is also to be designed, it is critical to identify and target the right sources where the data needs to be gathered and analysed as described before
Sometimes this phase is also compromised by the availability of the data, because of limitations set by many social media platforms as well as by technological limitations.
In cases where data accessibility is restricted and/or not available at all, it is recommended to consider using alternative sources and / or to shift the focus of the analysis.
To obtain raw data the main candidate sources are the following:
1. Social Networking Websites: they are a rich source of data for sentiment analysis applications.
2. Blogs are often made up of long texts describing an opinion with respect to a topic. They can also be crawled and annotated to create a sentiment dataset. Blogs tend to be structured narratives analysing the topic. They can be useful sources of data looking at different aspects of the given topic.
3. Microblogging (e.g. Twitter): Twitter API is a publicly available API that allows to download tweets based on a range of search criteria such as keyword-based-search, download timelines, download-tweet-threats, etc.
4. Discussion forums are portals where users discuss topics related to a central theme or initial question. Discussion forums usually arrange posts in a thread-like manner.
The list provided above is not exhaustive[2], but provides the list of the main data sources used in most of the social media analysis projects.</td></tr>
<tr><td rowspan="3">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>Health Canada</td><td></td><td>A wide range of sources (Twitter, GNIP, Tumblr, Facebook Pages, RSS feeds Blogs, News, Videos, Forums) has been taken in consideration.</td></tr>
<tr><td>NVWA</td><td></td><td>In order to overcome restrictions to the availability of data, the project has applied the approach to shift the focus of the analysis, e.g. using open data as a proxy of the more targeted closed data.</td></tr>
<tr><td>Other</td><td colspan="3">A. Joshi et al. Sentiment Resources: Lexicons and Datasets. Chapter in “A practical guide to sentiment analysis”, E. Cambria, D. Das, S. Bandyopadhyay and A. Feraco, Springer, 2017, pp 96-98.</td></tr>
</table>

<table>
<tr><th colspan="2">GUIDELINE 12</th></tr>
<tr><td>Area</td><td>Data Analytics</td></tr>
<tr><td>Topic</td><td>Planning – Data quality</td></tr>
<tr><td>Related business need</td><td>13. Availability of metadata (i.e., geospatial information) for the clusterisation of data would improve the quality of the analysis especially in terms of opinion trends monitoring</td></tr>
<tr><td>Description</td><td>Another important feature to consider while selecting and or generating the dataset to be analysed is related to the information embedded in the dataset itself.
The presence of metadata can provide the analysis with further insights and optimise the overall data analysis process.</td></tr>
</table>

[2] See D.1.1, “Social media use-cases, current usage and trends from public administrations”, par. 3.1, for a complete categorisation of social media.

<table>
<tr><td></td><td colspan="3">Metadata from social media spaces provides insights into geolocation, timestamp, connections between users, connections between users and hashtags, between users and @mentions, and between retweets/regrams/reblogs.
When working with metadata, data scientists can design their own tool by using the provided API key, which is available to developers for third party content creation. Of course this requires a working knowledge and the ability to program in the language that the API key requires.
Another possibility is to make use of pre-built tools for collecting metadata from social media sites. One such group who has created a plethora of tools for harvesting social media metadata is Digital Methods Initiative (DMI).</td></tr>
<tr><td rowspan="2">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>SENSEI</td><td></td><td>A scalable web-based software service mechanism offering SCRUD (search, create, read, update, delete) services over conversational data (in spoken or textual form) and associated metadata (annotations) has been developed within the project.</td></tr>
<tr><td>Other</td><td colspan="3">H. R. Gerber, T. L. Lynch. Into the Meta: Research Methods for Moving Beyond Social Media Surfacing. Tech Trends, May 2017, Volume 61, Issue 3, pp 263–272.</td></tr>
</table>

<table>
<tr><th colspan="2">GUIDELINE 13</th></tr>
<tr><td>Area</td><td>Data Analytics</td></tr>
<tr><td>Topic</td><td>Planning – Selection of the right technology</td></tr>
<tr><td>Related business need</td><td>14. Availability of software tools or services with adequate social media analytics capabilities</td></tr>
<tr><td>Description</td><td>The rising interest towards social media, made it possible that there is a vast offer of opinion mining tools skewed towards sentiment analysis of customer’s feedback about products and services. Such tools range from simple open-source tools to libraries, multi-function commercial toolkits and platforms. According to the scope of the analysis, the person in charge of the strategy selection, must perform a so called make or buy analysis, where the trade-off between developing the software tool in-house and make use of software tools available off-the-shelf is analysed.
To simplify the decision, below it is provided a description of the main tool categories available for data analysis purposes:
- Scientific programming tools: in order to provide support to sourcing, searching and analysing text popular scientific tools and analytics tools have been enhanced. These include: R, used for statistical programming, MATLAB, used for numeric scientific programming, and MATHLAB, used for symbolic scientific programming.
In addition to this, other programming languages such as Python can be used for (natural) language detection, title and content extraction, query matching and it can also be trained to perform sentiment analysis. Another language based on an open-source project that analyses big data and retrieves information that is relevant to the users is Apache UIMA (Unstructured Information Management Applications).
- Business toolkits: These toolkits are commercial suites of tools that allow users to source, search and analyse text for a range of commercial and not commercial purposes. An example of such toolkit is SAS sentiment analysis manager which can be used for scraping content sources, including mainstream Web sites and social media outlets, as well as internal organisational text sources, and creates reports describing the expressed feelings of consumers, customers and competitors in real-time.
RapidMiner is another toolkit that offers both an open source Community Edition license under the GNU AGPL and an Enterprise Edition offered under commercial license. It provides data mining and machine learning procedures including data loading and transformation, data processing, modeling, evaluation, and deployment. Another example is Lexalitycs that provides a commercial sentiment analysis engine for many OEM and direct customers as well as SPSS from IBM that is one of the most used programs for statistical analysis in social science.</td></tr>
</table>

<table>
<tr><td></td><td colspan="3">
- Social media monitoring tools: are sentiment analysis tools for tracking and measuring what are people saying about a company or its product or services or any topics across the social media landscape. Some examples of these tools are: Social Mention, which provides social media alerts similarly to Google Alerts; Amplified analytics which focuses on product reviews and marketing information; Lithium Social Media Monitoring; and Trackur, which is an online reputation monitoring tool that tracks what is said on the Internet. Google itself also provides a series of free tools. Further to these, it is worth mentioning Coosto, a single unique solution able to realise a social media monitoring, a management for the information needed, with the possibility to engage people involved. Coosto is used for different reasons; for instance, its ability to guarantee a control and measure over social media content is really appreciated by users. Another example of integrated platform is provided by the SENSEI platform which integrates several software modules such as: Syntactic parser, semantic parser, synopsis generator, sentiment analyser, abstractive cluster labeller, computation platform, co-reference resolver, website parser, discourse parser, agreement predictor, mood predictor, repository, repository tools, comment clustering and summarization, event/sentiment detector, social media eval prototype and ACOF (Agent Conversation Observation Form) tool (for more info please refer to the D1.1, “Social media use-cases current usage and trends”[3]).
- Text analysis tools: are tools for NLP and text analysis. There is a large number of freely available tools produced by academic groups and non-governmental organizations, for sourcing, searching and analysing opinions: Stanford NLP group tools and LingPipe, a suite of Java libraries for the linguistic analysis of human language, are some of them.
- Data visualisation tools provides business intelligence capabilities and allow different types of users to gain insights from the big data. Users can perform exploratory analysis through interactive user interfaces available on the majority of devices. Some example of BI visual dashboards are Qlik, Tableau and SAS Visual Analytics.
</td></tr>
<tr><td rowspan="3">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>ALL</td><td>various</td><td></td></tr>
<tr><td>NVWA</td><td></td><td>NVWA selected a tool able to produce query and reports, with a real time personalized targeting. About data collected, there is a high quality offered for data visualization (many layouts available) and a data mining capacity, able to guarantee the extraction of specific topics and news. Coosto is able to guarantee also geo location data, and the Authority has been using this information into its analyses.</td></tr>
<tr><td>Other</td><td colspan="3">B. Batrinca, P. Treleaven. Social Media Analytics: a survey of techniques, tools and platforms. AI and Society, Vol. 30, February 2015, Springer.
Feraco A., "How Visualisation and Interaction can optimize the cognitive processes towards Big Data", 4th "Encyclopaedia of Information Science and Technology-4th ed.", IGI Global, in February 2017.
Cambria E., Dipankar D., Sivaji B., Feraco A., Editors, “A practical Guide to Sentiment Analysis”, Socio Affective Computing 5, Springer.</td></tr>
</table>

GUIDELINE 14	
Area	Data Analytics
Topic	Planning – Selecting the right technology
Related business need	15. Significant computational capacity is needed to support the execution of simultaneous analysis on numerous users and in different languages
Description	Execution of simultaneous analysis on a vast amount of user-generated text and news content online implicates the content has high degree of disorder and diversity. In such respect NLP, computational linguistics and text analytics are deployed to identify and extract subjective information from the source text.

[3] See Annex 4, SENSEI Initiative sheet

<table>
<tr><td></td><td colspan="3">Automated sentiment analysis uses elements from machine learning such as latent semantic analysis, support vector machines, bag – of – words model and semantic orientation.
These techniques employ three broad areas:
- Computational Statistics: it refers to computationally intensive statistical methods including resampling methods, Markov chain Monte Carlo methods, local regression, kernel density estimation and principal component analysis.
- Machine learning: it applies when a system is capable of the autonomous acquisition and integration of knowledge learnt from experience, analytical observation and/or similar activities. These sub-symbolic systems further subdivide into:
o Supervised learning such as regression trees, discriminant function analysis, support vector machines.
o Unsupervised learning such as Self-Organising Maps (SOM), K-Means.
Machine learning aims to solve the problem of having huge amount of data with many variables and in commonly used in areas such as pattern recognition, financial algorithms, energy forecasting, biology etc.
- Complexity science: it is represented by the development of complex simulation models of difficult-to-predict systems derived from statistical physics, information theory and nonlinear dynamics.
The above techniques are deployed in two main ways:
- Data mining: knowledge discovery that extracts hidden patterns from huge quantity of data, using sophisticated differential equations, heuristics, statistical discriminators, and artificial intelligence machine learning techniques.
- Simulation modelling: simulation – based analysis that tests hypotheses.</td></tr>
<tr><td rowspan="2">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>NOMAD</td><td>  </td><td></td></tr>
<tr><td></td><td>STEP4YOUTH</td><td>  </td><td>The STEP Social Media monitoring tool is built on top of a set of independent services that are deployed over Docker containers that is a software “packaging” technology that facilitates deployment and testing. The tool consists of the following modules – services:
o StreamManager: this tool supports the continuous monitoring of five social media streams: Twitter, Facebook, Flickr, Google+ and YouTube to collect content relevant to a set of user – selected keywords, user accounts of locations, by using the corresponding APIs that are provided by each platform;
o Solr: This is an open source enterprise search platform built on top of Apache Lucene. It is used primarily for full text indexing of the collected social media items and the retrieval of content based on free text queries;
o MongoDB: this is an Open-Source, document database used to store the data collected from the StreamManager, such as Items, Media Items, Web Pages, Users, etc;
o Redis: This is an open source, in memory data structure store, used as a database, cache and message broker. It is primarily used as a publish/subscribe service to enable the different components of the tool to communicate with each other;
o Graylog: This is an open source log aggregator and management service, which aggregates and maintains the logs produced by the rest of the services.</td></tr>
</table>

Other	B. Batrinca, P. Treleaven. Social Media Analytics: a survey of techniques, tools and platforms. AI and Society, Vol. 30, February 2015, Springer.		

GUIDELINE 17			
Area	Data Analysis		
Topic	Planning – Selecting the right technology		
Related business need	18. Significant computational capacity is needed to support the execution of simultaneous analysis on numerous users and in different languages.		
Description	Public Administration could leverage on existing and reusable Social Media analysis IT platforms and/or OSS modules made available by previous initiatives. This approach could ensure to: - leverage on existing use-cases implemented by other PAs as well as to compare results of similar initiatives; - reduce costs of the implementation; - anticipate the start-up of an initiative, by using existing solutions/configurations. In any case a comprehensive and computational environment/facility will need the following attributes: - Data: o Data Scraping – the ability through easily programmable APIs to scrape any type of social media; o Data streaming – to access and combine real-time feeds and archived data for analytics; o Data storage – a major facility for storing principal data sources and for archiving data collected for the analysis and past references o Data security – the data stored needs to be protected to avoid any loss or theft; o Interface programmable – Data analysts need access to simple APIs to manage available data sources that may not be automatically collected. - Analytics: o Analytics dashboards – non programming interfaces are needed to grant access to what might be referred to as 'deep' access to 'raw' data o Programmable analytics – programming interfaces are also required so users can deploy advanced data mining and computer simulation using MATLAB, Python and Java; o Stream Processing – the computing environment needs to support analytics on streamed real-time data feeds o High – performance computing – the environment needs also to support non – programming interfaces to MapReduce/Hadoop, NoSQL databases and Grids of processors; o Decentralized analytics – in casa social media data need to be combined with highly sensitive and valuable proprietary data held by governments, financial institutions, retailers and other commercial organisations, then the environment needs in future to support decentralized analytics across distributed data sources and in a highly protected way.		
Reference initiatives	**Name**	**Nationality**	**How was the practice implemented?**
	NOMAD		
	HEALTH FOR CANADA		The Social Media Analysis tool used by Health Canada is a platform called Nexalogy, provided by the Canadian Environics Research Group. Nexalogy is a cloud service that optimize the computational power needed locally to analyse social

			media data. It helps discovering insights on social networks. Its main functionalities are: • Continuous retrieval and analyses of social media data (Twitter, GNIP, Tumblr, Facebook Pages, RSS feeds Blogs, News, Videos, Forums), using a powerful analytics engine; • Provision different visualisations: o Timeline o Top Content (most shared links) o Top terms (Hashtags and words) o Lexical Map of the top words and co-words and words clusters o Topics Map of 26 human aspects o Actors by Query (showing interaction between actors and queries) o Top Actors (Active and Mentioned) o Actor Interactions o Retweet Statistics • Capacity to create email alerts; • Capacity to create multiple personalized filters. The tool also offers geography and demographics features.
Other	N.A.		

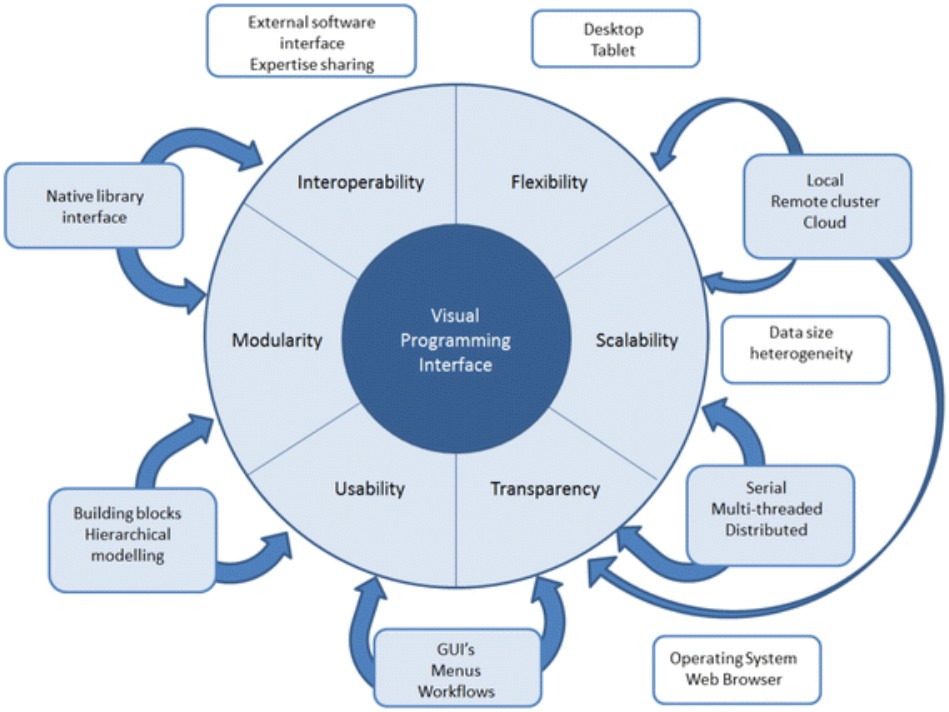
5.3. TECHNOLOGY, TOOLS AND CAPABILITIES

GUIDELINE 15			
Area	Technology tools and capabilities		
Topic	Execution – Software Maintenance		
Related business need	19. Ensure continuous maintenance of the SW platform that crawls data coming from proprietary external data sources (i.e. LinkedIn, Facebook, etc.) 23. Need for tools, either COTS or autonomously developed, customizable in accordance to the evolving requirements of the analysis.		
Description	Social Medias sometimes have a quick timeframe in changing their APIs and policies to share data. In this context, Social Media Analysis initiative could have some issues when the selected Social Media APIs and data are not stable and changes often. This event in fact lead to a continuous maintenance process in order to ensure the correct usability and correctness of the Social Media analysis, bringing increasing IT cost as well as risk of interruption of the data analysis processes. In order to overcome, or at list monitor, this challenge, the Public Administration that wants to keep its initiative up and running could: - base the analysis on Social Medias or on a set of data which are not affected by the changes of those data sharing policies (i.e. avoiding personal or sensible data etc, using aggregated data …); - implement control mechanism to ensure that as soon as any changes in the APIs is to be implemented, the IT platform will be not affected and will change accordingly. Moreover, in order to avoid an increase of IT costs, the Public Administration should have internally the IT staff with adequate competencies to regularly maintain the platform.		
	Name	**Nationality**	**How was the practice implemented?**

Reference initiatives	NOMAD		The NOMAD project was based on a custom IT Platform implemented internally thus having all the IT competencies to address the APIs changes.
	STEP4YOUTH		The STEP social media monitoring tool include the following technical features: - Creating collections from social media platforms and RSS feeds by entering keywords and/or user accounts of interest; - Collection of content in the form of items (posts made in social platforms, e.g. tweets, Facebook and YouTube videos, etc.); - Collection of contributors of social media content; - Detection of dominant topics and languages in each collection; - Filtering of items in a collection based on language, platform (Facebook, Twitter, YouTube, etc.), publication date (since-until) and originality (original content or shared); - Sorting of items according to publication time (recency), popularity and relevance; - Analytics over collections with a number of visualisation widgets.
Other	N.A.		

GUIDELINE 16	
Area	Technology tools and capabilities
Topic	Execution – Software Look&Feel
Related business need	20. Develop user-friendly yet effective tools that may cope with the limited familiarity with IT technology in public administrations, especially at local level.
Description	IT platforms, mainly the OSS and/or custom-made ones, have to be developed in order to be easily usable and configurable without any additional IT intervention. The need of a easy-to-use look and feel is in line with the need to overcome also both business needs #1 and #12 linked to the cultural barriers and low level of digital skills of Public Administrations. In order to implement a proper look and feel, the overall framework must be designed ad hoc to satisfy also non-technical people and with different cultural backgrounds. To do so, several disciplines may support the overall software architecture and/or integration. One of the most commonly used is the Visual Programming (VP). A visual Programming language is a medium for implementing computer programs that makes use of graphical operators and elements rather than textual ones. VP is different from GUI. A GUI aids users executing programs via visual menu items in contrast to command-line (i.e. terminal) text scripting. In general, GUI menus are premade and users cannot create new programs or combine menu functions within the GUI. Conversely, a VP language has the same power as a textual programming language or a library, if it features the same functional elements (e.g. data structures and methods); therefore new algorithms and programs can be designed and compiled within a VP, and VP can even be used to implement GUIs. Although in principle VP can be used to create algorithms starting from the lowest hierarchy of programming language elements, in practice, VP is employed for creating higher-level applications using libraries. This facilitates developers' work when a large amount of coding (and redundant coding) is required. The main elements of a VP languages are: building blocks, block engines, block connectors, and meta-blocks.

By definition, VP facilitates various aspects of the process of software design and brings closer the user and the developer.
Development policies for VP should follow the same prerogatives. One way to achieve this can be through the adoption of the Agile methodology [81], which implements code and develops user interfaces at the same time, with an adaptive strategy of real-time planning. Agile is an efficient and face-to-face communication between developers, stakeholders, and final users, blending the characteristics of different disciplines directly into the final product.
Independently of the chosen the development methodology, a VP framework, in this case, for Social Media Analytics, should meet the following requirements: flexibility, scalability, transparency, usability, modularity, and interoperability. Each of these characteristics affects either a user's or a developer's needs.
For instance, flexibility of a VP product concerns its capacity to be workable on different devices such as desktops vs. tablets, or feasible for local/server/cloud installations (which is different from being installable on different operating systems or being usable). Scalability refers to the capacity of the software to scale up with data size increase, by featuring different solutions, such as switching to multithreaded mode or moving analyses from a localhost to the cloud.
Transparency can be associated with multiplatform (any operating system) or web-browser-based implementations. The VP framework and its products must be transparent not only at the operating system level but also at the hosting level (i.e. local host vs. cloud).
Within a developer's VP environment, usability can mean a well-designed set of building blocks and functions. These are also related to modularity, where the availability of a native generic template programming library can make a difference. Finally, interoperability can be divided into three levels. The first one is software interoperability, which ensures the possibility of using external pieces of software or libraries. The second level is semantic interoperability, which is the ability to exchange data with unambiguous, shared meaning. The third level is expertise interoperability, connecting the stakeholders together efficiently by using an appropriate communication infrastructure, such as a users' forum or a developers' space, like GitHub. A summary of the requirements for a VP framework is provided by the following image:



The use of VP is highly recommended when aiming to achieve a very highly customised "look & feel" level, although it needs a very deep knowledge of the various programming languages.

Name	Nationality	How was the practice implemented?

Reference initiatives	STEP4YOUTH		The Step platform specifically aims at a user-friendly interaction with all stakeholders, making use of several visual features. For example: - the front page of the platform is set so to visually allure visitors to interaction; - from the front page, users are led to pages where, again, extensive use of visual features is made (e.g. maps to locate the various relevant initiatives, timelines where users can place events and thereby leave their footprint, etc.); - visual representations of results are meant to enable policy makers to interpret the results of their campaigns, view patterns and spot trends, facilitating decision making. The initiative has also made particular attention to the inclusion of young people with fewer opportunities (because of social/economic obstacles, cultural differences, etc.), including dedicated workshops for those who are not familiar with the use of digital tools, info kiosks, and special dissemination material.
Other	Milicchio F., Rose R. Bian J., Min. J., and Prosperi M. "Visual Programming for next-generation sequencing data analytics" in BioData Mining, 2016.		

GUIDELINE 17	
Area	Technology tools and capabilities
Topic	Execution – Openness and Reusability
Related business need	21. The technology and tools supporting the analysis must be open source and reusable. 23. In case of use of social media analysis infrastructure, this should satisfy at least the following requirements: - clear SLAs in terms of uptime / performance - adequate information security - customization possibilities - software compatible with the IT architecture of the PA
Description	Public Administration could leverage on existing and reusable Social Media analysis IT platforms and/or OSS modules made available by previous initiatives. This approach could ensure to: - leverage on existing use-cases implemented by other PAs as well as to compare results of similar initiatives; - reduce costs of the implementation; - anticipate the start-up of an initiative, by using existing solutions/configurations. In the NOMAD initiatives, the platform implemented is a multiple domain platform able to scan several data sources, and hence developed to be re-used by further following initiatives. Another example of open and reusable software architecture is represented by the SocialSTORM platform developed by the UCL in UK. SocialSTORM is a cloud-based platform which facilitates the acquisition of text-based data from online sources such as Twitter, Facebook, respected blogs, RSS media and 'official' news; a 'central - hub' for social media analytics. The system includes facilities to upload and run Java-coded simulation models to analyze the aggregated data; which may comprise UCL's social data and/or users' own proprietary data. There is also connectivity to the ATRADE platform which provides further quantitative finance and economic data. The basic idea behind the SocialSTORM platform is that each external feed has a dedicated connectivity engine (API) and this streams data to the message bus, which handles internal communication, analytics and storage. The main components of the platforms are:

<table>
<tr><td></td><td colspan="3">Connectivity engines: Various connectivity modules communicate with the external data sources, including Twitter & Facebook's APIs, financial blogs and various RSS news feeds; and are being continually expanded to incorporate new social media sources. Data are fed into SocialSTORM in real-time and include a random sample of all public updates from Twitter, as well as filtered data streams selected from a rich dictionary of stock symbols, currencies and other economic keywords; providing gigabytes of text-based data every day.
Messaging Bus: This serves as the internal communication layer which accepts the incoming data streams(messages) from the various connectivity engines, parses these (from either JSON or XML format) and writes the various data to the appropriate tables of the main database
Data Warehouse: this is home to terabytes of text based entries which are accompanied by various types of metadata to expand the potential avenues of research. Entries are organized by source and accurately time-stamped with the time of publication, as well as being tagged with topics for easy retrieval by simulation models.
Simulation Manager: this terminal provides the external API for clients to interact with the data for the purposes of analysis, including a web-based GUI via which users can select various filters to apply to the datasets before uploading a Java-coded simulation model to perform the desired analysis on the data.</td></tr>
<tr><td rowspan="3">Reference initiatives</td><td>Name</td><td>Nationality</td><td>How was the practice implemented?</td></tr>
<tr><td>NOMAD</td><td></td><td>The individual tools can be further developed, customised and reused in a follow-up initiative. Each service within the platform, is owned by the department participating in the project that has developed it. Apart from the visualisation services, all partners have made their code available as open source.</td></tr>
<tr><td>DG COMM INITIATIVES</td><td></td><td>The tool that supported the 4 DG COMM initiatives projects is ANACONDA. Based on open source technologies, Anaconda is the leading open data science platform powered by Python and R programming languages for large-scale data processing, predictive analytics and scientific computing. The open source version of Anaconda is a high performance distribution of Python and R and includes over 100 of the most popular Python, R and Scala packages for data science.</td></tr>
<tr><td>Other</td><td colspan="3">Wood R., Zheludev I., and Treleaven P. "Mining Social Data with UCL's SocialSTORM Platform. " from http://www.academia.edu/4144373/Mining_Social_Data_with_UCLs_Social_STORM_Platform.</td></tr>
</table>

5.4. INITIATIVES' KPIS AND KSFS

The investment decision of social media technologies and applications in government must be part of the organization's enterprise architecture planning effort (Dadashzadeh, 2010).

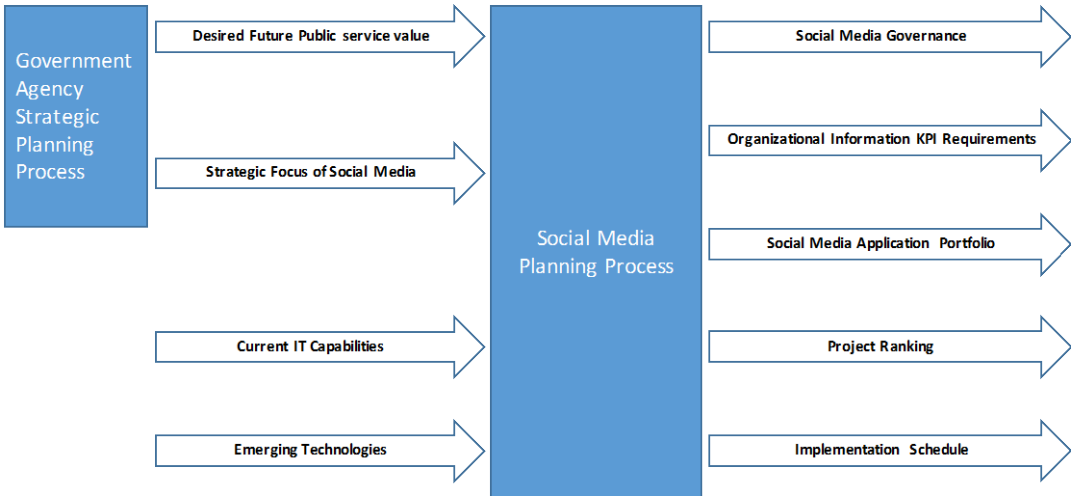


Figure 2 - Input/Output Model of IT Planning for Social Media in Government.

The desired public service value to be created and the role of social media in realizing it must be guided by four principles:

1. Outcomes-based focus: aim to generate tangible improvements in the social and economic conditions of citizens
2. Balanced to ensure fairness: serving the common good by providing access for all citizens
3. Engagement to Co-Produce Public Value: Engage, educate, but also help citizens improve their own quality of life by tapping into their experience
4. Improving Government Accountability: increasing transparency in reporting cost-effectiveness of initiatives and providing citizens the opportunity to "talk back" when governments fail to deliver expected public value.

Social Media enables internal communication and collaboration, exchange information with the public, and keep pace with fast moving events in '***real-time"***. Given the four inputs listed in the Figure 2, the social media process must produce five outputs:

Each social media initiative should serve a well-defined, mission oriented purpose, and have the potential to improve agency's communication effort. It is then imperative to define key performance indicators (KPI) that define what success looks like and how it will be measured. After all, although social media can uncover cost savings, its strength lies in increasing citizen engagement to harness the collective ingenuity of the public to coproduce tangible improvements for the common good.

Similar to any other IT investment, social media investments need to be planned and the required concomitant organizational change to culture, people, structure, and processes be managed for effective results. In such vision, a strategic planning approach becomes necessary for effectively leveraging social technology in government.

While setting the KPI, the following issues related to social media need to be addressed within the strategy definition:

- Ensuring information disseminated through social media is consistently available;
- Making information available through social media available in other formats for those who lack equal access due to infrastructure, ability, language, or literacy;
- Maintaining consistency of access for government agencies and for members of the public;
- Archiving information disseminated through social media for permanent access and retrieval;
- Preventing release of sensitive or secret information;
- Fostering transparency and accountability, through which government is open and transparent regarding its operations to build trust and foster accountability;
- Ensuring the security of personally identifiable information;
- Maintaining security of user information;
- Providing a continuously updated data.gov registry, with an historical index that shows current and past data availability;
- Ensuring that third-party social media technology providers (e.g., Twitter, Facebook, YouTube, Second Life) adhere to government privacy, security, and accuracy policies and requirements;
- Ensuring that individual-government transactions that transpire through social media technologies are confidential, private, and/or secure as required by federal laws and policies;
- Ensuring continuity of service, especially when technologies sunset.
- Ensuring that mashups and other forms of data integration—an increasing activity due to data availability via data.gov—do not lead to user profile development that invades privacy or otherwise compromises individuals, national security, or agency data security;
- Monitoring the storage of government information when held offsite through cloud computing services. Allowing private companies to maintain potentially sensitive government data raises

enormous questions of data retrieval, accuracy, and permanence, as well as opens up significant opportunities for misuse of data by providers or attempts by other governments to access the data based on the geographic location of the server farms where the data are maintained; and

- Ensuring that social media technologies are not the only means of getting a response from an agency.

This list is by no means exhaustive, and each type of social media technology raises its own specific set of specific KPI.

On the other hand, KPIs can also be set quantitatively in respect of the goals the overall analysis has per target. Such KPIs can be determined by:

- Number of young people involved in the various initiatives
- Number of policy makers involved in the various initiatives
- Number of decision – making procedures piloted through the initiative
- Number of additional entities (both private and public) interested in the adoption of the technology / app developed during the initiative
- Number of targeted people reached

When setting the Key Success Factors (KSFs), conceptual achievements should be determined. I.e.

- Identifying clear steps for the implementation of the initiatives with specific deliverables for each step;
- Ability to access also closed sources to deepen the analysis and identifying also existing correlation with not p[publicly available data
- Setting up collaboration / partnerships with relevant stakeholders that enable the optimization of the knowledge base needed (IT experts, different set of expertise under the same roof for the project scope, etc. etc.)
- Ability to predict and forecast the outcome of specific events (i.e. see SENSEI project and its BREXIT prediction)

6. CONCLUSIONS

Despite the widespread and increasing availability of online opinions, needs and preferences expressed by citizens on the various Social Media, European Public Administrations are, so far, still not able to consume this unstructured and dispersed knowledge in order to extract meaningful knowledge and use it as input to policy and decision making.

Within this context, the **ISA^2 Action 2016-04 – Participatory Knowledge for supporting Decision-Making**, particularly the project stream on "**Leveraging Social Media full potential to increase citizen engagement and participation in public administrations' decision-making processes"**, aims at:

- on one hand, identifying existing replicable Social Media analysis initiatives (and reusable tools, technologies and techniques used for data processing) and,
- on the other hand, at providing a set of relevant guidelines for those Public Administration that are willing to start from scratch using Social Media analysis and related supporting IT tools.

To this extent, the analysis of identified Social Media initiatives across EU, has led to the identification of several elements that can be considered good practices, together with a number of key success factors and potential business needs of EU Public Administrations in this domain. Such elements mainly refer to key areas supporting the startup and implementation of Social Media Analysis projects supporting policy- and decision-making[4] that drove the delivery of *"Guidelines for the Implementation of Social Media Analysis initiatives"* to be used for the potential implementation of future initiatives and pilots in the Social Media Analysis field. In particular, the main areas for which those guidelines have been developed are related to:

- **How to design of the overall vision and strategy** underlying the Social Media Analysis initiative, in view of the final goal to be achieved in terms of support to policy- or decision-making. This area includes how goals and purposes have been communicated and disseminated;

[4] See D01.01, "Social media use-cases, current usage and trends from public administrations", p. 57

- **How to start-up and implement a specific initiative**, in terms of clear steps for the effective and successful start-up and execution;
- **Approaches and methodologies for data analytics**, referring to the methodology and tools utilized to collect/cleanse/analyse/process information available on Social Media in order to obtain relevant result for policy- and/or decision-making purposes;
- **Reference technology and tools**, in terms of main features and functionalities of the technology application developed (or purchased) to support the initiative, with special regard to the possibility of re-usability, openness and modularity of the technology and platforms currently used by the identified initiatives.

The guidelines described in this document have been structured in order to provide a clear link to the related business need(s) expressed by the stakeholders of the analysed initiatives:

- have a clear framework in terms of **data privacy** regulations and policies, in order to ensure access to **reliable data** and **metadata**, allowing proper clusterisation according to the requested analysis to be performed;
- raise the **awareness and engagement of public administrations** potentially interested in social media analysis initiatives;
- **improve the digital skills**, specifically related to Social Media and Data Analytics, of personnel directly or potentially involved in such initiatives;
- exploit **open and re-usable social media analysis infrastructures** providing **adequate IT tools** and **customisable data analytics features** and ensuring a **reliable service level**.

In conclusion, the present document represents a first reference for European Public Administrations that are willing to start leveraging Social Media full potential to increase citizen engagement and participation in public administrations' decision-making processes but are still on the starting blocks and have not a clear idea on what and how to do it.

6.1. WHAT'S NEXT

In accordance with the ISA^2 Programme mission and general principle to support, promote and encourage re-usability and avoidance of duplication of solutions and standards, the next step of this project should envisage the following main steps:

- **Milestone 1:** delivery of a feasibility study aimed at scoping and designing the implementation of a Big Data pilot to support the specific "Social Media Analysis" use case in the context of the **ISA^2 action 2016.03 - Big Data for Public Administrations: Big Data Test Infrastructure.** In this context, the study must include an assessment of the level of re-usability, interoperability of the IT tools supporting the identified existing Social Media initiatives.
- **Milestone 2:** execution of the selected Pilot in collaboration with a Member State and in the context of specific policy domain or internal decision-making process. This step will include the dissemination and communication activities related to the pilot execution.
- **Milestone 3:** delivery of a feasibility study aimed at the generalisation of the identified Social Media Analysis IT platform / tools in order to make it available to European Public Administration. This step will include the analysis of the governance model for the use of the IT platform / tools.
- **Milestone 4:** generalisation of the identified Social Media Analysis platform / IT tools and availability on JoinUP and in the context of the Big Data Test Infrastructure (when available).
- **Milestone 5:** dissemination and communication activities related to the availability to Member States of this new IT tool and capability provided by the ISA^2 Programme.