D2.4 Initial evaluation and report of updated requirements

Project acronym:	**SENSE4US**
Project full title:	**Data Insights for Policy Makers and Citizens**
Grant agreement no.:	**611242**
Responsible:	**Timo Wandhöfer (GESIS)** Ruth Fox (Hansard Society)[1]
Contributors:	Rasa Uzdavinyte (Gov2u) Benjamin Zapilko (GESIS)
Document Reference:	**D2.4**
Dissemination Level:	**PU**
Version:	**1.0**
Date:	**26/10/2015**

[1] Ruth Fox took over as deliverable lead in October 2015, because Timo Wandhöfer left GESIS.

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

History

Version	*Date*	*Modification reason*	*Modified by*
0.1	30/09/2015	Initial draft	Timo Wandhöfer
0.2	04/10/2015	Updated draft	Benjamin Zapilko
0.3	11/10/2015	Updated draft	Rasa Uzdavintye
0.4	22/10/2015	Updated draft	Ruth Fox
0.5	22/10/2015	Quality check	Steve Taylor
1.0	26/10/2015	Final reviewed deliverable	Ruth Fox, Benjamin Zapilko

Table of contents

List of figures

List of tables

List of abbreviations/ glossary

Abbreviation	Description
PM	Project Month – PM1 is the first month of the project, etc.
WP	Workpackage
DoW	Description of Work
PR	Press Release
Research Partners	The partners responsible for the research components of the project: U Koblenz (WP4), Open University (WP5), U Stockholm (WP6)
UI	User Interface
ULEV	Ultra Low Emission Vehicles
End user partners	The partners within the project responsible for end user engagement in WP2: GESIS, Hansard Society, Gov2u
End users (or simply “users”)	The actual end users consulted by the project’s end user partners.

Executive summary

This deliverable sets out the method and results of the pre-evaluation testing and end user evaluation of the early versions of the Sense4us prototype toolbox and its components, exploring how the end-user partners tested initial requirements at the regional, national and EU policy-making levels, and formulated new or modified existing requirements based on the feedback received.

Between PM13 and PM24, three prototype versions of the toolbox were rolled-out by the research partners for end-user testing. This deliverable is therefore structured to reflect that multi-stage, iterative testing process.

Part 1: explores the method by which the end user partners sought to evaluate each prototype version, exploring the benefits and challenges arising throughout the process.

Parts 2-4: set out the findings from the testing of Toolbox versions 1-3 respectively.

Part 5: sets out two case study examples of how the Sense4us tool might be used in real world research and policy environments – one in Parliament and one in government – based on interaction with an end user who undertook sustained testing of Toolkit version 2.

Part 6 highlights the topline results – exploring the key findings and emerging feedback themes and how these have been addressed.

1. Introduction

This deliverable summarises the activities and progress of the SENSE4US project's end user engagement in its second year. It follows from the initial requirements deliverable D2.1, but this deliverable is different in scope and format from D2.1. This is because the end user engagement has taken a different form in year 2. During year 1, the requirements were based on a survey and explanation of concepts to prospective end users, and the result of this was D2.1. During year 2, the project has completed three prototypes, building on each other, and these were the subject of the evaluation engagement. Hence this deliverable concentrates on feedback from the end users regarding these prototypes and the resulting requirements arising from this feedback. In addition, this deliverable indicates which tools and which features were seen as useful, and which were not.

This deliverable sets out the method and results of the pre-evaluation testing and end user evaluation of the early versions of the Sense4us prototype toolbox and its components, exploring how the end-user partners tested initial requirements at the regional, national and EU policy-making levels, and formulated new or modified existing requirements based on the feedback received.

Between PM13 and PM24, three prototype versions of the toolbox were rolled-out by the research partners for end-user testing. This deliverable is therefore structured to reflect that multi-stage, iterative testing process.

The rest of this section explores the method by which the end user partners sought to evaluate each prototype version, exploring the benefits and challenges arising throughout the process.

Parts 2-4: set out the findings from the testing of Toolbox versions 1-3 respectively.

Part 5: sets out two case study examples of how the Sense4us tool might be used in real world research and policy environments – one in Parliament and one in government – based on interaction with an end user who undertook sustained testing of Toolkit version 2.

Part 6 highlights the topline results – exploring the key findings and emerging feedback themes and how these have been addressed.

1.1 Related Deliverables

Deliverable	*Citation*	*Month*	*Short description*	*Input/ Output*
D2.1 Assessment of the proposed end user requirements	[D2.1]	9	D2.1 addresses the feedback gathered from all three policy making levels, regarding the project's aim to create a toolbox to facilitate the policy making process at the resource and policy modelling levels.	*D2.4 input:* general end user requirements
D3.1 First Functional requirements description	[D3.1]	12	D3.1 describes the process and results of work to find the initial functional requirements of the Sense4Us toolkit. Specifically, it describes the steps taken to transform the user requirements from D2.1 into the functional	*D2.4 input:* functional requirements

			specification that is the overall goal of this deliverable.	
D2.3 Defining policy scenarios	[D2.3]	12	The purpose of D2.3 is to define scenarios and policy issues at each of the different policy-making levels, which can be used to drive and evaluate the ICT research carried out by the consortium.	*D2.4 input:* scenarios and policy issues
D3.3 Second Functional requirements description	[D3.3]	26[2]	Updated version of D3.1.	*D2.4 output:* updated list of requirements

Table 1: Overview of related deliverables

[2] Due to the two months' delay in this deliverable, D 3.3 is also delayed by two months.

1.2 Summary of Initial Requirements

Below are the priority requirements that emerged out of the end user research from PM1-PM12 as reported in D2.1. For reference a full list of functional requirements can be found in D3.3.

Provenance

- Difficulty of identifying 'trusted' sources
- Trustworthiness largely based on researchers judgement
- Be able to assess/compare data provenance
- Provide supporting information about the sources used: e.g. peer review / citations / availability of data & clarity of authorship etc.

Understanding/ Usability

- Be transparent about the methodology underpinning the tool → build trust/confidence
- It needs to be 'easy' to use – end-users won't spend a long time 'learning' how to use it (training / toolkits)
- Provide accessible forms of presentation – e.g. visualisation

Search functionality

- Broad range of sources and formats
- Multiple variables / customisability
- Cross-disciplinary search across multiple sources & formats (time-saving)
- Capacity to save search history

Finding related information

- Multiple information sources - internal and external data sources
- Combined sources – e.g. media/social media with official legislative documents
- Cross-border information / global data
- Time-frame comparisons (length of life)

Filtration and presentation of results

- Filtering and structure of results – e.g. multiple filters – linked to provenance
- 'Show' and 'hide'
- Visualisation of datasets / results (particularly for non-expert / non-technical users – e.g. regarding statistics)
- Result summaries – overview enabling links / next step searches to be determined quickly

The development process is such that not all functional requirements were addressed in the first versions of the toolkit.

1.3 Conceptual Overview

End user evaluation has been an iterative process as the three versions of the Sense4us toolbox have been rolled-out for testing purposes over time.

The system architecture design was created taking into account the functional requirements derived from end user research in the first year of the project. An iterative cycle of 'implement-demonstrate-feedback' then followed. Early feedback from testing of Toolbox v1, combined with more advanced development of functional requirements, were incorporated and rolled out in Toolbox v2, a process that was then repeated for Toolbox v3 creating a rolling process of development as illustrated in Figure 1 below.

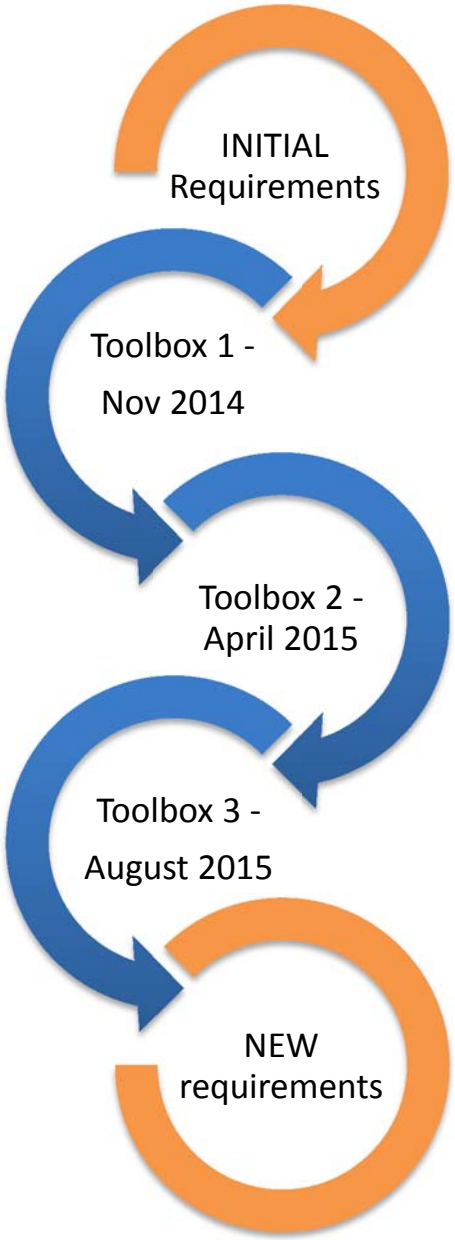

Figure 1: Conceptual overview initial evaluation, toolbox versions 1-3

Table 2 below shows the stages at which new tool components came on stream for evaluation purposes in this three stage testing cycle.

Toolbox's components	*Toolbox v1 (Nov 2014)*	*Toolbox v2 (April 2015)*	*Toolbox v3 (August 2015)*
Search Twitter:	✓	✓	✓
Upload documents:	✓	✓	✓
Main themes:	✓	✓	✓
Sentiment:	✓	✓	✓
Theme overlap:	✓	✓	✓
Locations:	✓	✓	✓
LOD search:		✓	✓
LOD surrounding:			✓
SentiCircles:			✓

Table 2: Functional components available for testing by toolbox version / date

1.4 Challenges

The end user partners faced a number of very practical challenges – both logistical and conceptual in nature - when testing the various versions of the Toolbox, as a result of which it was necessary to adopt an iterative approach that was responsive to local circumstances rather than a common model deployed across all three policy-making levels**.**

1.4.1 Timeframe / events

A significant challenge - particularly in the UK – has been the way in which events, primarily the Westminster general election and its aftermath in May 2015 – effected the Hansard Society's ability to test the Toolbox versions as they came on stream with a variety of end users, including those who participated in our interviews and focus group research in the first year of the project. The UK general election – and the uncertainty arising - was highlighted as a potential high risk factor in the project proposal and Description of Work, and indeed proved to be problematic. As detailed in section 1.7, the project responded creatively to the problem posed, but the longest election period in the UK since 1928 inevitably hampered end user engagement. The situation was further exacerbated by the fact that Parliament went into recess again soon after re--convening following the election, and only returned fully with effect from week commencing 12 October.

Research partners made every effort to deliver the prototype versions mindful of the timetable sensitivities for end user partners, but inevitably there could not be a complete alignment of technological developments for roll out in the most optimal weeks and months for end user engagement, not least as a sustained period of pre-evaluation testing by end

user partners themselves was necessary. In order to demonstrate and evaluate the Toolbox versions, end user partners need to have a full understanding of how to use the Toolbox in each new form so external testing could not begin as soon as the prototype became available – there was a time-lag as end users familiarised themselves with it and indeed tested it for early identifiable problems.

1.4.2 Engaging non-technical audiences with prototypes

Most policy makers – the core user audience for the Sense4us tool – have a non-technical background. As such, their understanding of concepts such as sentiment analysis, theme overlap, policy modelling and simulation is limited at best, and they struggle to understand the connectivity structures and relationships between search terms.

Ultimately, end users care how the tool will help them do their job – they want to see instant and compelling search results that they can immediately appreciate the benefit of in terms of the workflow and research challenges they face. However, early prototype models by their very nature cannot provide such instant, compelling outputs as the functionality and UI are at an early stage of development. The first two versions of the toolbox were particularly difficult to test with end users because of the early functional and User Interface (UI) constraints – early results perceived to be 'disappointing' in terms of the output risked having a detrimental impact on end user engagement and feedback provision. At the prototype stage there is a gap between the 'actual' and the 'possible' in terms of the search outputs and therefore expectations of end users have to be carefully managed. As such, testing of the earliest versions of the toolbox with some end users was more appropriate than with others and necessarily influenced how we demonstrated the toolbox to different audiences. For example, for some end users such as MPs the tool was demonstrated via pre-prepared test cases that illustrated the potential of the toolbox; but parliamentary clerks, librarians and policy advisers with more time available were targets for more sustained engagement including direct testing of the toolkit themselves. Looking ahead, this realistic constraint on end user engagement has informed how we have set about developing a series of worked policy examples that we can demonstrate to a number of end user audiences.

1.5 Method and Objectives

1.5.1 Pre-evaluation testing

All end user partners dedicated significant amounts of time after the launch of each Toolbox version to pre-evaluation testing in order to learn how to use the Toolbox, exploring both how to use the components separately, and in combination. During these periods of testing, gaps and glitches were identified so that the relevant research partner(s) could eliminate them before the toolbox was shown to end users. Drawing on the end user partners' own experience of the work of policy makers and the challenges they face, modification opportunities were also set out in order for work to begin on early improvements to functionality. Through this process, end user partners also identified and worked up small demonstration cases to be used as part of the end user engagement process.

1.5.2 Demonstration and feedback

All three end user partners had common objectives in terms of their approach. These can be summarised as:

- Obtaining feedback on individual components of the toolbox;
- Obtaining feedback on the various toolbox components chained together;
- Obtaining feedback on the toolbox's usability;

- Identifying the current practices of policy makers to better understand and explore the information challenges they face and how Sense4us might address these;
- Mapping information sources for case study examples;
- Investigating perceived/highlighted shortcomings in the versions of the toolbox;
- Translating that feedback into new or revised requirements to develop and improve Sense4us.

However, given the nature of local challenges and constraints, each end user partner adopted a different approach to achieve these objectives.

1.6 Regional level: State Parliament of North Rhine Westphalia (GESIS)

"Leibniz meets Parliament" is an event that frequently takes place within German parliaments. The objective is to present and discuss the current research of the Leibniz' institutions with decision-makers to make use of the research results in daily working situations. These events are key forums for demonstrating Sense4us to parliamentarians and their staff and seek feedback.

A few weeks before the event all members of the parliament receive a cover letter including a list of the topics offered by the researchers of the participating Leibniz institutes.

Timo Wandhöfer (GESIS) showcased Sense4us at three events:

i) 3 December 2014 – 'Using online data for the decision-making process' - State Parliament of North Rhine Westphalia, Dusseldorf - Toolbox v1

ii) 21-22 April 2015 – 'Using online data for the decision-making process' - Bundestag, Berlin – Toolbox v2

iii) 3 September 2015 – 'Policy making with Open Data and Twitter' - State Parliament of North Rhine Westphalia, Dusseldorf - Toolbox v3

In advance of the meetings with MPs, research was conducted to ascertain information about the issues they were interested in, the debates they had been contributing to, and what their potential research information requirements might be. This intelligence then informed the development of personalised use case scenarios for presentation to them, and the collection of data required in order to run the scenarios. A demonstration of elements of the toolbox – e.g. a Twitter search, or theme analysis of a document – was then conducted either directly, or via presentational slides (based on pre-prepared data and screenshots of the prototype in action). Throughout the demonstration discussions took place with the interviewee to explore their response to the presentation, whether they considered the output data useful (and if not, what improvements might make it so), what metrics would help them interpret the results, in what kind of scenarios they could imagine utilising the tool, and what modifications would improve the benefit and utility of the toolbox for them both in terms of ease of use and presentation of outputs, and possible development of further functionality.

The discussions were necessarily flexible – no set script was adhered to, the approach was defined by the individual concerned – as it is important that the end users direct the feedback rather than the project partners.

The following table shows those MPs interviewed as well as key identification information about them. The last two columns show the key search terms and policy documents that were used to pre-prepare test cases in order to demonstrate the potential of Sense4us in an engaging and relevant way for each MP. A demonstration case encompassed either social media or policy documents, and, where time permitted, both forms of search.

Table 3 is organized by the MPs „profiles“. But not all of the MPs within the table participated in the interviews. In some cases only the clerks provided the feedback. In total 12 MPs comprised our core feedback group, with several of them responding to at least two different versions of the prototype. Additionally, interviews and demonstrations were also held with 13 parliamentary clerks via these events. Several more interviews / demonstrations were prepared but the MPs had to cancel at short notice. This is a perennial problem with end user engagement which, whilst frustrating, has to be expected.

MP	*Party*	*Website*[3]	*Twitter name*	*Constituency*	*Search terms*	*Policy Document*	*Toolbox version tested*
Hans Feuß	Social democrats (SPD)	http://www.hans-feuss.de/	@hansfeuss	Gütersloh II	#inklusion, #schule, #bildung, #bildungssystem, #sport, #kultur, #medien	UN convention for disability rights[4]	V1
Arne Moritz	Christian democrats (CDU)	http://arnemoritz.de/	---	Solingen I	#solingen, #bauen, #wohnen, #stadtentwicklung, #verkehr, #kommunalpolitik, #selbstverwaltung, #petition, #mittelstand		V1
Torsten Sommer	Pirates	http://www.torsten-sommer.de/	https://twitter.com/tosopiratas	Dortmund	#dortmund, #flüchtling AND arbeitsverbot, #gesundheit, #technologiepark, #hartz-4, #bürgermeisterabwahl		V1 & V3
Thomas Nückel	Liberal (FDP)	http://www.thomas-nueckel.de/	https://twitter.com/ThomasNueckel	Herne	#kultur, #medien, #meinungsfreiheit, #informationsfreiheit, #presse, #landesmediengesetz, #grunderwerbssteuer, #medienordnung		V1 & V3
Andreas Nick	CDU	http://www.dr-andreas-nick.de/	https://twitter.com/DrAndreasNick	Montabaur (number 205)	#bundestag, #digitalisierung, #flüchtlinge	Die Woche im Bundestag vom 23. bis 27. März 2015[5]	V2

[3] Retrieved on 25/09/2015.

[4] Cp. URL: http://www.behindertenrechtskonvention.info/ (Retrieved on 25/09/2015).

[5] Cp. URL: http://www.dr-andreas-nick.de/aktuell/newsletter-aktuell.html (Retrieved on 25/09/2015).

Kai Whittaker		http://www.whittaker.de/	https://twitter.com/Kai_Whittaker	Rastatt (number 273)	#griechenland, #bundestag, #mindestlohn	Rede im BT zum Mindestlohn [6]	V2
Thomas Jarzombek		http://jarzombek.de/	https://twitter.com/tj_tweets	Düsseldorf	#dortmund, #flüchtling AND arbeitsverbot, #gesundheit, #technologiepark, #hartz-4, #bürgermeisterabwahl	Request of the government: Human and decentralized accommodation of refugees.[7]	V2
Gerold Reichenbach	SPD			Groß-Gerau	#kultur, #medien, #meinungsfreiheit, #informationsfreiheit, #presse, #landesmediengesetz, #grunderwerbssteuer, #medienordnung	Cost Savings in Policing[8]	V2
Patrick Schnieder				Bitburg (number 203)	#berlin, #breitbandausbau, #mobilität		V2
Inge Blask	Social Democrats (SPD)	http://www.inge-blask.de/	https://twitter.com/ingeblask	Balve, Hemer, Menden, Neuenrade und Plettenberg (number 122)	#Klimaschutz, #Umwelt, #Landwirtschaft, #Naturschutz, #Verbraucherschutz, #Wirtschaft, #Energie, #Industrie, #Mittelst, #Handwerk		V3
René Schneider	Social Democrat (SPD)	http://www.reneschneider.de/	https://twitter.com/schneider_rene	Wesel II	#Datenschutz, #Digitalisierung, #Internet, #Medien, #Politik, #Big Data, #Soziales, #Telekom		V3
Robert Stein	Christian Democrats (CDU)	http://robert-stein.nrw/	https://twitter.com/RobertStein_MdL	Hamm (number 118)	#Haushalt, #Finanzen, #Digitalisierung, #Startups, #cnetz		V3

[6] CP. URL: http://www.whittaker.de/bundestagsreden/2630/ (Retrieved on 25/09/2015).

[7] CP. URL: http://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMD16-4164.pdf (Retrieved on 25/09/2015).

[8] CP. URL: http://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMD16-7320.pdf (Retrieved on 25/09/2015)

Daniel Schwerd	Pirates	http://www.daniel-schwerd.de/	https://twitter.com/netnrd	Köln	#Urheberrecht, #Privatsphäre, #Transparenz, #Familie, #Bildung, #energie		V3

Table 3: Profiles of participating MPs' in 'Leibnitz meets Parliament' events

1.7 National level: Westminster, UK (Hansard Society)

As noted in section 1.4.1, end user engagement in the UK has been particularly difficult as a result of the lengthy election period, swiftly followed by the summer recess and then the party conference recess period. In the few weeks that Parliament sat in June and early July it was extremely difficult to arrange meetings with end user partners to test Toolbox v2 in both Parliament and government departments because they were absorbed with new work priorities arising from the setting up of the new Parliament and government and, as is usual in the immediate aftermath of an election, many parliamentary and departmental staff were moving on to new roles within their institutions as part of the natural churn and professional development of staff who rarely stay in roles for more than a few years at a time.

The election and its aftermath had been identified as an unavoidable risk at the start of the project, but with the agreement of project partners a creative solution to the difficulty was pursued. An application was made to the House of Commons Management Board in March 2015 for a parliamentary committee clerk to be allocated a 'dissolution placement' for the duration of the election period in order to spend time at the Hansard Society testing the Sense4us toolbox and assisting our research more generally by contributing to our ideas for use case scenarios and the electric cars test case. A committee clerk is an example of a researcher who is targeted with investigating a policy subject area, or to evaluate the pros and cons of a particular piece of draft legislation, so this was a good fit with the project. It was agreed that the clerk should have an interest in information management, research and issues such as open data, social media etc, and that the placement should be mutually beneficial, contributing to the clerk's professional development as well as aiding the Society and the wider Sense4us project. The Management Board duly agreed to the proposal – an innovation for the use of parliamentary staff during an election period at Westminster – and invited expressions of interest from staff. A number of clerks responded and Edward Faulkener, clerk to the Political and Constitutional Affairs Committee in the House of Commons, was duly confirmed.

He subsequently spent time over a five week period at the Hansard Society helping to test the Sense4us prototype. An induction day was spent with project partners at IT Innovation, Southampton, in order to introduce him to the project and prototype Toolbox (version 2), he was set up with a log-in and password, and then returned to the Society to conduct extensive tests on the system. In addition to pre-evaluation testing by Society staff of all three toolbox versions, this additional period of sustained testing of Toolbox v2 encompassed document search, social media search and sentiment analysis, linked open data search, and the user interface. A number of use case scenarios were developed and tested, a detailed policy scenario – including identification of key stakeholders and policy makers, relevant documents, data and social media search terms – was worked up in relation to the issue of Ultra Low Emission Vehicles, and the toolbox's utility considered in the context of potential use by a parliamentary select committee.

1.8 Supra-national level: European Union (Gov2u)

At the EU level, pre-evaluation testing by the end user partner took place on Toolbox 1 and 2, with end user demonstration and feedback focused on Toolbox 3 due to concerns about resourcing and the development of compelling demonstration scenarios from the early Toolbox versions, not least because of language constraints.

At the time of writing, following a period of pre-evaluation testing, direct demonstration of Toolbox 3 to end user partners is in its early stages. Three end users from the wider EU policy making environment have been chosen for interviews – the potential end users were picked not only from the EU institutions, but also member state government levels and stakeholder organisations that work actively in the EU policy making process. Such a broad approach to end user engagement will help ensure the potential of the tool is optimised and meets the needs of a range of policy makers working in different information and research environments.

The following table shows those interviewed as well as key identification information about them: the environment in which they work and their policy areas of interest.

	Position/organisation	***Policy domains***
End user 1	EU consultancy, Legal and Public affairs consultant	ICT, Digital Single Market, Internal Market
End user 2	EU Parliament, MEP adviser	Internal Market and Consumer protection, ICT, Employment and Social Affairs, Foreign policies
End user 3	Attaché, Member state government representative at the European Council	Environment

Table 4: End users for EU-level engagement

As with the interviews conducted at the regional level, demonstration cases drawing on the end users' areas of personal interest were developed encompassing, in these instances, either social media or linked open data searches. Using the Toolbox, relevant searches were run in order to collate pre-prepared data for presentation during the interviews, and in two of the interviews additional live tweet searches were conducted by the interviewee leading to a generation of data which then informed discussion of the functionality and results.

2. Toolbox v1 (Launched November 2014)

2.1 Regional level: State Parliament of North Rhine Westphalia - Demonstration Cases

Although GESIS carried out the end user engagement for Toolbox v1, there was co-ordination with the other end user partners regarding the focus of the demonstration and feedback sessions. Five core elements were explored:

1. Presenting the concept and basic functionality of the toolbox
2. Querying Twitter with search terms plus theme / sentiment analysis
3. Uploading of one of the MPs' documents (policy document plus topic analysis)
4. Theme overlap of the MPs' policy document and a twitter collection
5. Mentioned locations

2.1.1 Feedback

The participants were attracted to the concept of the toolbox and its potential overall, raised no concerns regarding data protection issues, and did not comment adversely on the interface directions 'Source', 'Analysis' and 'Trust'. Overall, however, a significant concern was the need to modify the tool such that its functionality would work with the German language.

To aid usability, some end users suggested the tool be available in two settings: 'light' (normal) and 'advanced' mode. This would give the user greater personal control over the level of search being conducted and information provided as a result. A user may not wish to use all the functionality. Giving them the option to choose the level at which they deploy the tool would also have the advantage of simplifying the user interface thereby aiding navigability for less technically adept users.

It was noted that the more pages of search results that were produced (via either document or twitter searches) the greater the clutter on the interface and the more difficult it was for the user to maintain an overview of the original search they had undertaken. A delete function was therefore noted as a necessity. The use of icons and labelling (e.g. Twitter, sentiment, file document etc) were also hard to decode if a search results list became long and there was a desire for a clear contrast in the colours used to visualise sentiment analysis.

Participants liked the way in which users could organise and easily observe the history of their searches, and the resulting analysis, via the 'My projects' section of the toolbox. However, the option to share the results with relevant people would augment the utility of this functionality. They also noted the importance of what was described as the 'delta' – namely what changes had taken place in respect of their searches for particular projects since they last logged in. And all participants mentioned the value inherent in being able to monitor key words over a period of time.

Some of the functionality supporting Twitter search was not working properly – e.g. the provision of a google map linked to tweet outputs - and was deemed of limited value if such searches could not be location based.

In terms of sentiment, interviewees broadly welcomed the opportunity to see the distribution of tweets and sentiment over time and how this changed within and between days, as illustrated in Figure 2 below. However, having highlighted peaks and troughs in tweets and sentiment, users wanted to dig down further to better understand what accounted for the pattern.

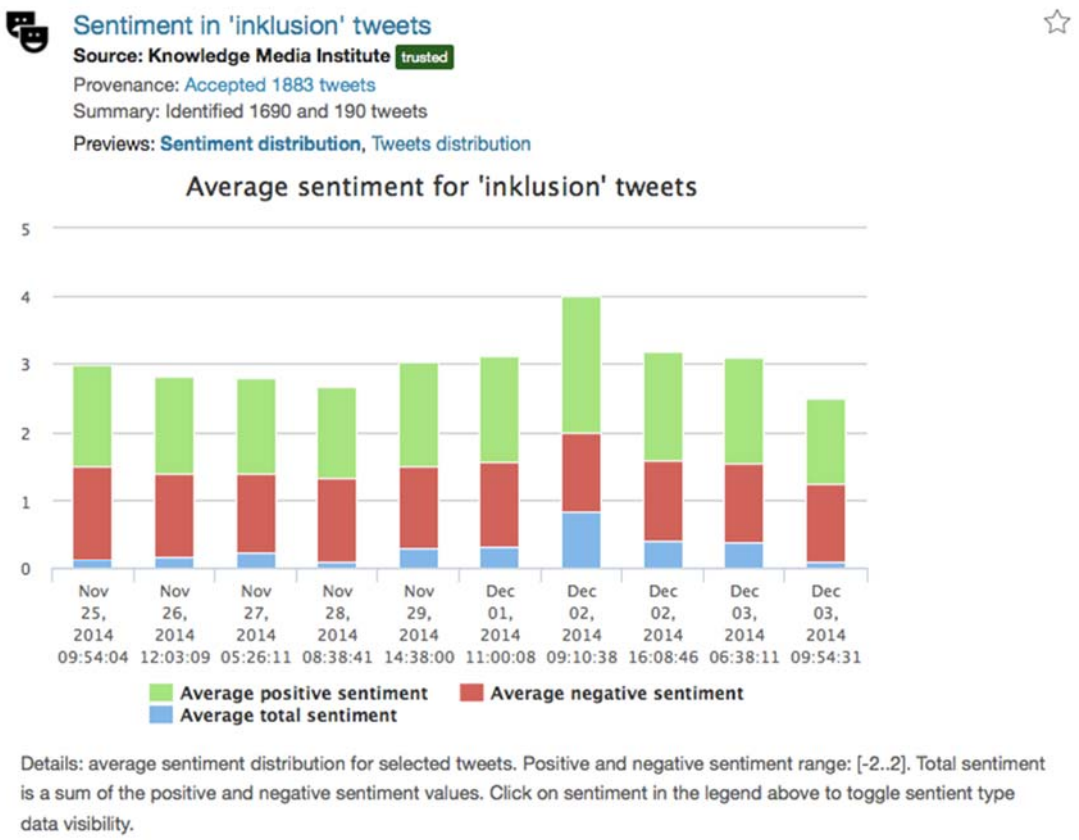

Figure 2: Toolbox1 – Screenshot Average sentiment for 'inklusion' tweets (integrating the handicapped into mainstream education)

If the user clicked on a sub collection of negative or positive tweets an ordered list of tweets emerged and they wanted to know the ranking order of them. Here, the wording on the user interface had to match the visual outputs so that positive came first on both or vice versa (but not positive/negative on one but negative/positive on the other).

If lists become long and complex, users suggested they would want to know about sub-groups such as the most influential users concerned – for example, domain experts such as journalists – and therefore a filtering function would be useful.

Users also wanted greater flexibility in the selection of topics when looking at the themes of tweets. The toolbox provided for a maximum of four themes but this was deemed insufficient. And when those themes emerged users wanted to drill down in more detail to visualise the influence of certain themes. For example, when a main theme analysis of 'inklusion' (integrating the handicapped into mainstream education) tweets was run, the topics referred to tagespiegel (a website) or tatort (a television programme). MPs felt it would be helpful to visualise the correlation of sentiment on particular topics with the time the programme was broadcast.

Users liked the concept of theme overlap analysis, but felt that the data that emerged – e.g. see Figure 3 - was insufficiently transparent to be useful – it was not clear how many paragraphs within a document the search was based on and what percentage of the document it concerned in contrast to the proportion of tweets. The headline theme overlap also caused confusion as it occurred only in relation to twitter when it was deemed potentially useful and interesting in connection with other potential search combinations.

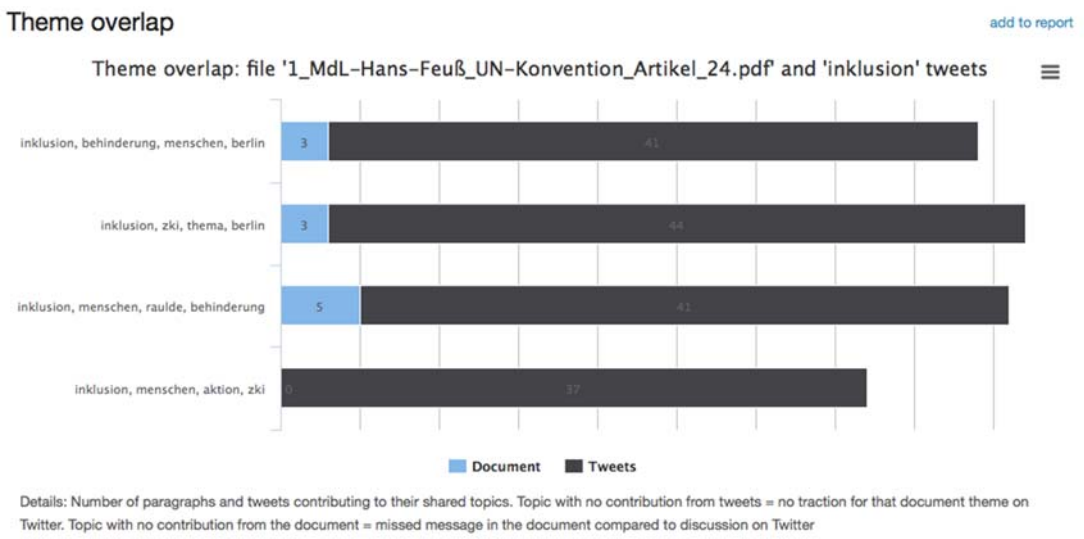

Figure 3: Toolbox1: Screenshot Theme overlap

Feedback was also received, particularly from the MP, Torsten Summer, regarding possible scenarios in which the toolkit might be used and how the results might usefully be visualised for non-technical end users such as parliamentarians. He pointed, by way of example, to the Federal Institute for Employment's[9] tool to visualise employment rates across Germany. Here, as Figure 4 and Figure 5 show, a dashboard allows for different results arising from multiple search functions to be displayed alongside each other.

Figure 4: Federal Institute for Employment's tool to visualise employment rates across Germany

[9] http://www.arbeitsagentur.de/web/content/EN/index.htm]

Key facts and figures about employment in Germany are displayed alongside a news feed focused on particular keywords, and relevant reports in the policy field. A coloured heat map that distinguishes clearly between a number of chosen variables – in this case unemployment rates at state level – was deemed helpful. Being able to visualise Sense4us search data in similar ways would be useful to those in the policy field, enriching understanding of existing data.

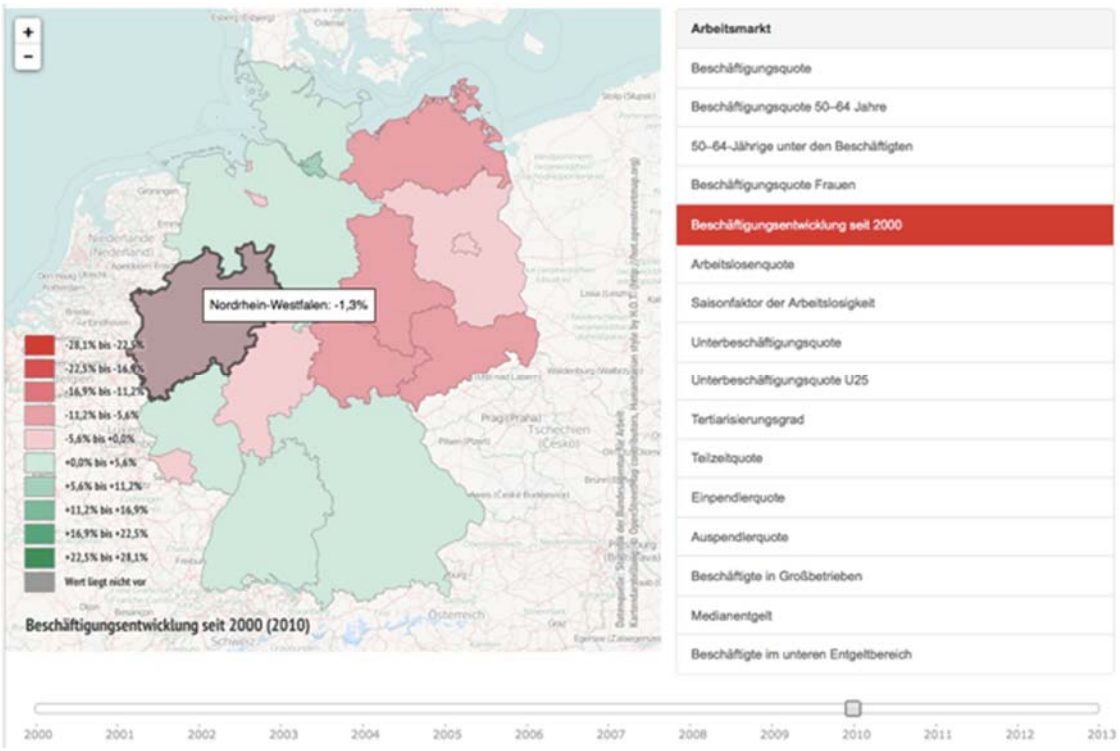

Figure 5: Federal Institute for Employment's tool to visualise employment rates across Germany - detail

3. Toolbox v2 (launched April 2015)

3.1 Regional level: State Parliament of North Rhine Westphalia - Demonstration Cases

For this version of the toolbox, three key functions were demonstrated and tested:

a) Tweets analysis:

Step 1: Start a search

- Input: search term (e.g. electric car)
- Output: A list of tweets, accompanied with information regarding the number of tweets, the number of unique users and the time frame of the collection

Step 2: Run sentiment analysis

- Input: a collection of tweets
- Output: sentiment of tweets (positive vs. negative) over time

Step 3: Run topic analysis

- Input: a collection of tweets

  Output: a set of four keywords that appear often within the tweets.

b) PDF document analysis:

Step 1: Upload a PDF document

- Input: Drag & Drop or select from files
- Output: a list with paragraphs from the document

Each paragraph has a sentiment

c) Tweets and PDF document comparison:

Step 1: Start a search

- Input: search term (e.g. electric car)
- Output: A list of tweets, accompanied with information regarding the number of tweets, the number of unique users and the time frame of the collection

Step 2: Upload a PDF document

- Input: Drag & Drop or select from files
- Output: a list with paragraphs from the document
  - Each paragraph has a sentiment

Step 3: Run topic overlap

- Input: select a twitter collection and PDF document

  Output: a list of keywords that frequently occur in both collections.

3.1.1 Feedback

Twitter search and sentiment

The sheer volume of tweets that emerged at the initial search stage made it difficult for users to get an overview of the outputs. Particular concern was expressed about the visualisation of

sentiment via 'smiley face' icons (see Figure 6 below). It was unclear what these meant – scales were deemed necessary for evaluation purposes. Users also questioned the basis for the sentiment – why 'fracking' was viewed positively, for example, was unclear. Users argued that perceptions might be rooted in political perspective and preferences that could change over time.

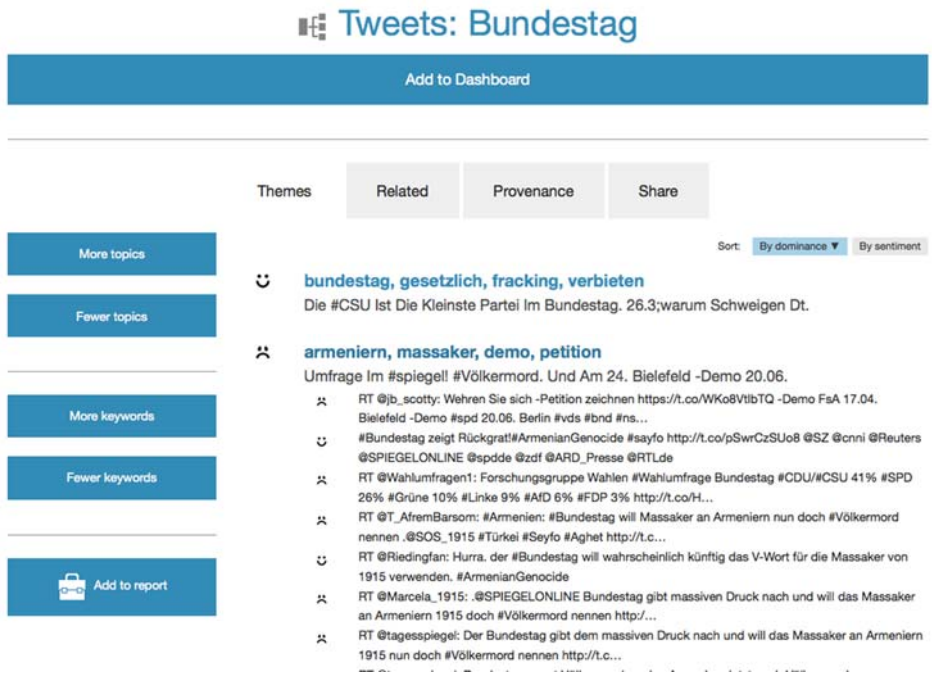

Figure 6: Toolbox2 – Screenshot Theme analysis for the collection of tweets for 'Bundestag'

An alternative demonstrator case looked at data privacy – from the state's perspective, data privacy, which it intended to pursue, was a positive issue. However, a politician from the opposition party might think very differently and view the topic negatively. A Twitter search for the topic "Betreuungsgeld" (care allowance) demonstrated the concerns in compelling terms (see **Figure 7**).

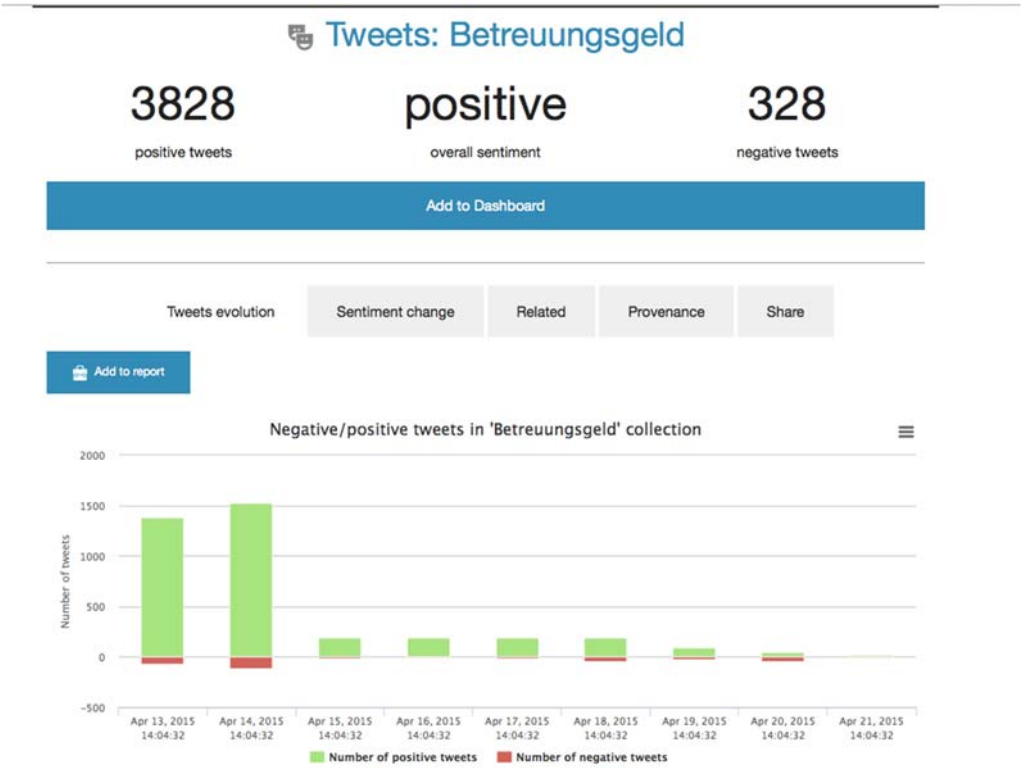

Figure 7: Toolbox2 – Screenshot Negative/positive tweets in 'Betreuungsgeld' collection

At a political and policy level, consideration was being given to abolishing it. How, then, were a large number of tweets emerging with positive sentiment? MPs wanted to know what the basis for this was - who are the opinion makers upon which the data draws.

Another demonstrator case, twitter sentiment in relation to the search term 'Dusseldorf', resulted in the visual output in Figure 8. Here, interviewees queried how and why the division of sentiment was so stark – what was the criterion as it was not immediately clear from the results. Once again, however, interviewees – albeit with one exception - welcomed the ability to see sentiment develop over time; this was thought likely to be helpful in order to detect outliers. Here users were interested in knowing when there was a departure from the trend, and if so, why.

And, as previously, they were keen to be able to dig deeper to better understand how and why peaks and troughs occur. Users wanted to know 'what happens next', 'where are the arguments', 'where can I go into the discussion?' Incorporating the number of tweets, the period of time for the search, and the number of users contained in the search output, were thought to be potentially useful visual enhancements.

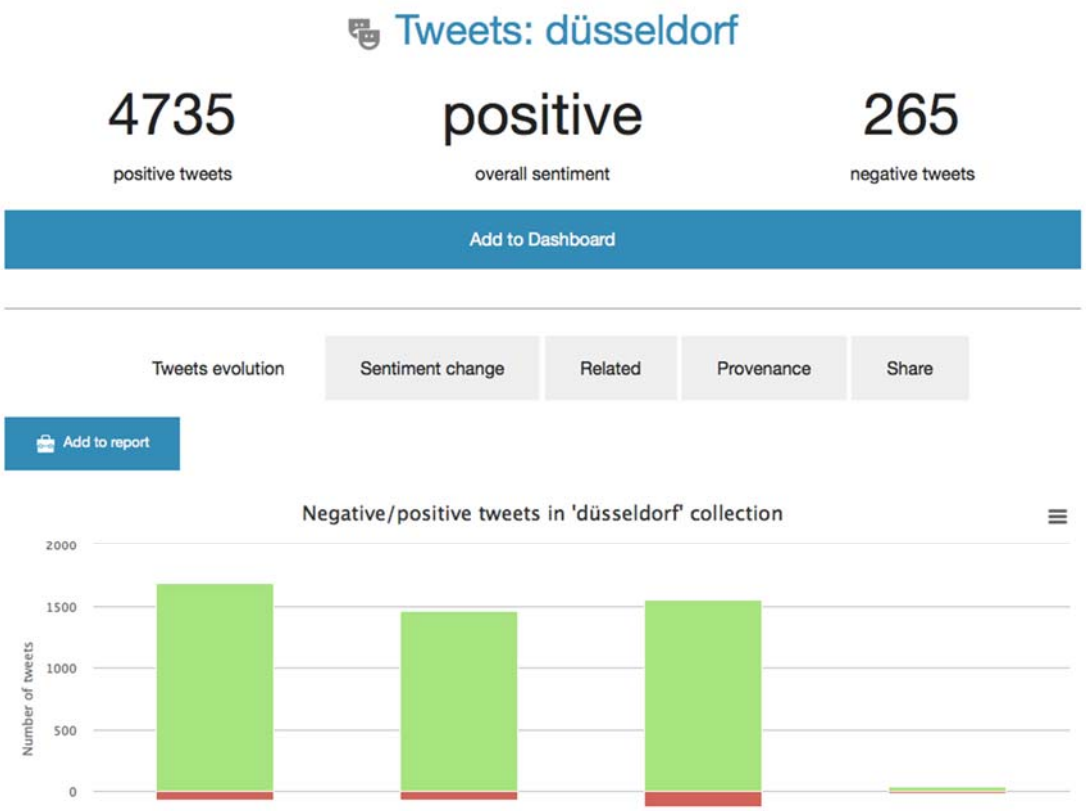

Figure 8: Toolbox2 – Screenshot Overall sentiment for twitter collection 'düsseldorf'

Overall, then, a key requirement was to be able to identify the Twitter actors involved in the search output and if possible order the tweets by number of followers or perceived impact. Additional supporting information about the topics and how the sentiment relating to them was assessed was also deemed necessary. The information in the provenance structure did not provide this – there needs to be more detail about the 'workings' by which the analysis is arrived at. Clear display of attachments to tweets – e.g. pdf files, website links, images - was also considered useful in terms the overall objective of being able to find related information.

Theme overlap: document upload & comparison with Tweets

The concept of 'main themes' was deemed confusing by some as no main topics are shown but rather most frequently used words / groups of words. Especially in combination with sentiments, this was considered confusing.

MPs compared their newsletters about current affairs in Berlin with tweets about the Bundestag. The output data, illustrated in **Figure 9**, was deemed potentially useful if MPs could regularly update it to match the production schedule of the newsletters. This would enable parliamentarians to check the extent to which the newsletter matched discussion of the same issues on Twitter and the areas and issues where overlap existed.

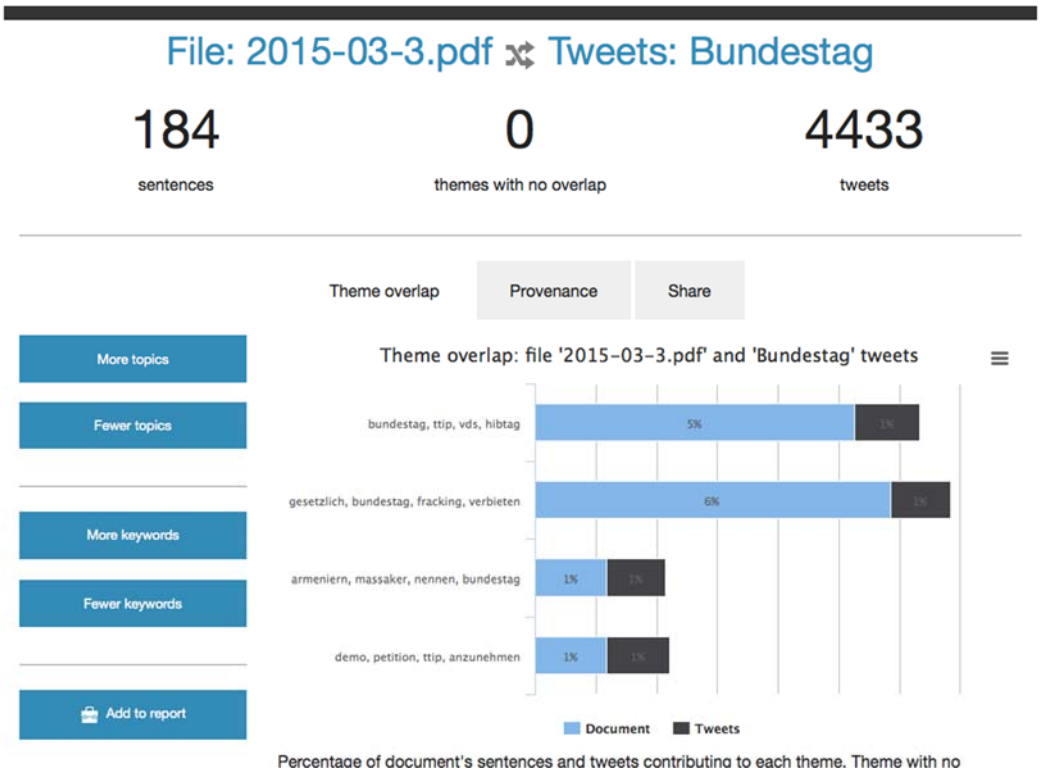

Figure 9: Toolbox2 – Screenshot Theme overlap PDF-document and tweets 'Bundestag'

An alternative demonstrator case explored the theme overlap between the search of a speech about the Bundestag and "Mindeslohn" (minimum wage) and comparison with the same on Twitter, as illustrated in Figure 10.

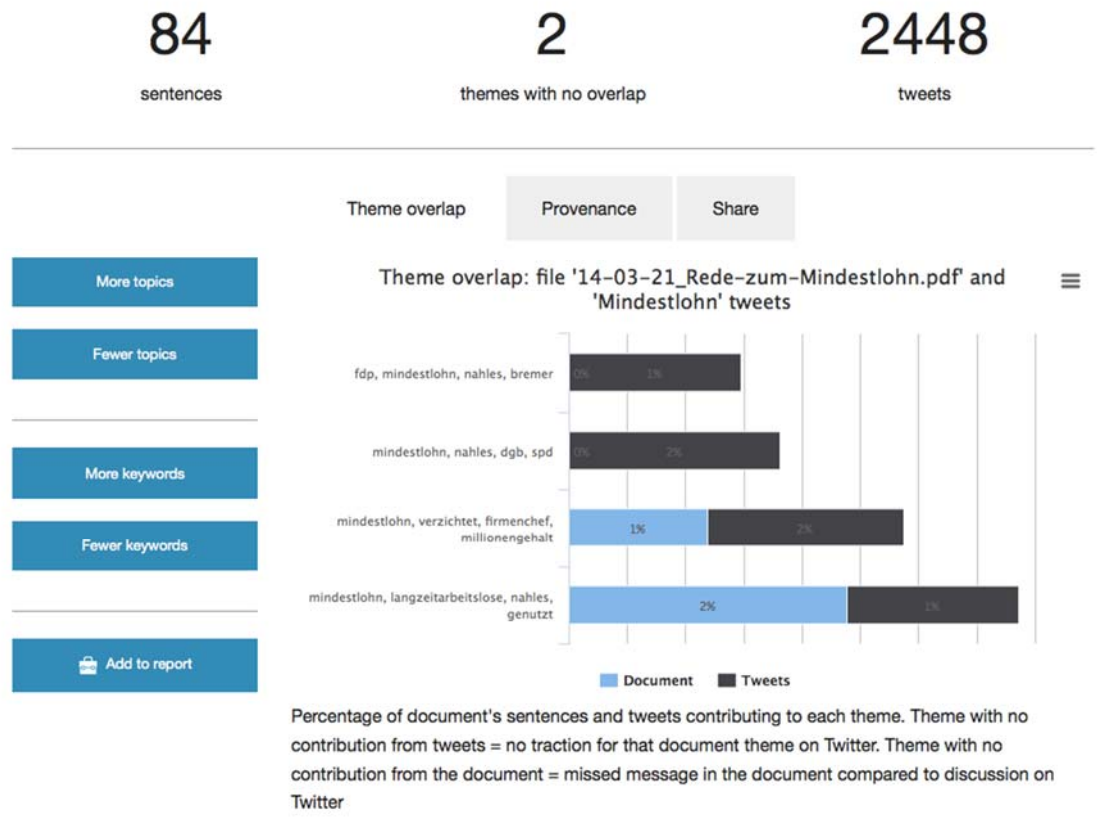

Figure 10: Toolbox2 – Screenshot Theme overlap PDF-document and 'Mindestlohn' tweets

Here, the numbers assigned to the bars were deemed meaningless – it was unclear what the percentages meant. More information was necessary in order for the graphic to convey meaning in accurate terms – for example, what are the crucial and essential tweets that overlap? Although the users claimed the results were 'interesting' they could not see how this functionality would be usefully applicable in their daily work. At one interview, it was suggested that there would be more value in being able to group and contact Twitter users.

It was also suggested that incorporation of data from Facebook in addition to Twitter, and on blogs instead of documents would be helpful. For example, the top 100 blogs by reputation could be included as they discuss relevant topics in more detail and a lot of the discussion that takes place online ultimately finds its way into traditional media. One user suggested it would also be useful, to contact users regarding their constituencies or topic related issues of interest.

3.2 National level: Westminster, UK

Several weeks of sustained testing of Toolbox v2 focused on issues relating to the functionality and interface and the overall Sense4us toolkit experience.

It should be noted that a number of the issues outlined below were also picked up by GESIS and Gov2u in their own pre-evaluation testing but are listed only once below to avoid unnecessary duplication in this report.

3.2.1 Feedback

Document uploading

Difficulties were experienced uploading certain types of document (using either the "Select File" or "Drag and Drop" functions). A red cross appeared on screen but no explanation was given for it. Users initially thought this was to be used to remove an item – it was not immediately clear that it was because there was a problem with the document upload.

Once the "process document" button was selected, the screen displayed a "processing, please wait" message but despite refreshing the document didn't appear. Users need "information feedback" and "error messages" to explain problems and highlight actions that need to be taken as they move through the various search stages.

Document display

An important problem was that the tool couldn't extract and search text from tables embedded within documents. This is a particular concern for government and parliamentary documents where tables with bullet point information are often inset into the wider text in order to highlight information. The search function also rejected short or single sentences, excluded bullet points and other short pieces of text – it was suggested that the user should be able to set the criteria for automatic rejection/acceptance of paragraphs of text. An 'accept all' button was also desired as on many occasions the user found himself wanting to accept many of the paragraphs that had been automatically rejected – it would have been easier to accept them all than reject the few he didn't want. Occasional problems were also experienced with English documents being listed in German and then resulting in no search information being displayed.

Tweets

Although a reject/delete option for Tweets had been introduced into this second version of the Toolbox, testing demonstrated that it did not work and needed refinement.

Social media search

It was suggested that it would be useful to be able to search for Tweets that include "at least 2 of these words" (an open search returns too many irrelevant results, but requiring several words returns too few – a happy medium is required). The ability to conduct future tracking/monitoring, was deemed a very helpful potential function. Thus might be done via an option to 'turn on' a search to be run 'live' over a period of time; it could be made a 'favourite' or added to the dashboard. Inclusion of a 'history' option would also help support this function.

As many of the tweets were not country based they were deemed of limited relevance to Parliament/Government in terms of an assessment of public attitudes. For example, unless it explicitly wants to explore an overseas issue, a parliamentary committee investigating ULEV policy will primarily want to know about UK based opinion and sentiment, not that of countries such as Germany.

It was also noted that there was a high level of commercial/advertisement tweets for which a delete/mark repeats option would be necessary.

Sentiment analysis

Concern was expressed that the positive/negative indicators for Tweets (for example in relation to the ULEV scenario) didn't seem to match a common sense reading of Tweets. It was suggested that it might be better to give a numerical indication of positivity/negativity as a happy/unhappy categorization was somewhat blunt, and not very helpful when many tweets may be neutral.

Extending the sentiment function to documents was also deemed helpful – for example, online comments received in response to public policy surveys and consultations could then be analysed. At present, for example, parliamentary committee staff have to perform an 'eyeball' assessment and analysis of comments received via online forums – if the volume is high this can be very difficult and time consuming. Sense4us tools could help address this and introduce more rigour into the process.

Insights – main themes

As with other users, the four words in a theme were not deemed particularly helpful and were not always interrelated in a helpful way. A most common words/statistics on word usage indicator was thought likely to be more helpful to users.

Linked open data search

Having failed initially to get this tool to work, it became clear that users need to give the website permission to run scripts or the search fields do not populate with text from the Linked open database. In Chrome it wasn't obvious what the problem was. In Opera the browser stated that the page was trying to do something and needed permission. This issue needs to be highlighted for the user so that they can take whatever action is necessary depending on the browser they are using.

It was recommended that users be able to drag and drop parts of the visual results map so that they could view the results more flexibly. The initial results were always shown from the minimum distance of 0 connections – it was suggested that it would be better to default to the middle distance or provide hits for all distances. It was also noted that having the type of link (eg. Wiki PageExternalLink) listed in full for each result could rapidly result in a lot of clutter. Abbreviating the links or colour coding them might be a better way of highlighting their provenance. An option to go to the webpage each link relates to was also recommended as a helpful improvement.

Overall, search results were found to be 'very hit and miss' – it took a lot of trial and error to get a return with any links, but on occasion did produce a very useful map. As an example of

where it was useful, an LOD search for “First-past-the-post voting → Electoral reform” returned a map of links with many relevant keywords and links to two organizations which work in the area of electoral data and policy, both of which produced information relevant to the subject and might be interested in a Committee’s work in this area.

4. Toolbox v3 (launched September 2015)

4.1 Regional level: State Parliament of North Rhine Westphalia – Demonstration cases

Tweets analysis

Step 1: Start a search

- Input: search term (e.g. electric car)
- Output: A list of tweets with information regarding its number, the time frame of the collection and the unique user

Step 2: Run sentiCircles

- Input: a collection of tweets

  Output: sentiment of tweets (positive vs. negative) over time

Open Data & Linked Open Data

A demonstration of the concept of LOD search was presented to interviewee's as the search functions were not fully operational at the time of the Leibnitz event.

4.1.1 Feedback

4.1.1.1 Open Data & Linked Open Data

The LOD search accepts two search terms in order to find links and relationships between information. So, for example, a search for 'solar' and 'NRW' (North Rhine Westphalia) may provide information about solar energy connected with the NRW region, for example local companies that are engaged in the process of solar energy.

The interviewees' saw lots of potential in the concept as it is not currently possible to run this kind of search query on the type of sources currently being used. However, to be effective and utilised, the tool will need to be developed to work across a wider range of Open Data sources. Wikipedia, the current database on which it draws, is regarded as a useful first step overview source but not somewhere policy-makers go for detailed information and analysis.

Another issue is that the component needs to accept German key words as search input and provide German documents as output.

4.1.1.2 SentiCircles analysis regarding topic-based search

The initial results using the new sentiment tool, OU's "SentiCircles", suggested a more favourable outcome than the previous tools used in Toolbox v1 and v2. Perceptions of positive and negative sentiment appeared better aligned to the nature of the public policy debate than with the previous sentiment analysis tool, and MPs who saw it were more comfortable with the messages it conveyed.

Figure 11 shows the SentiCircle results for the 'thematic' hashtag '#refugeeswelcome' in Germany.

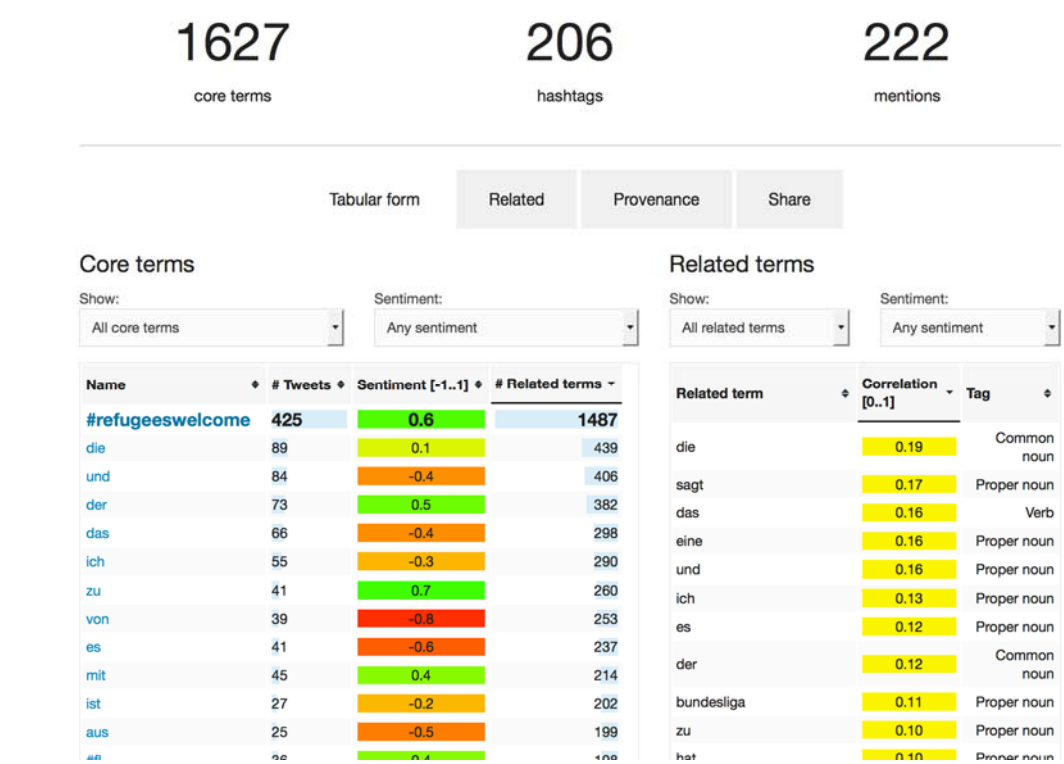

Figure 11: Toolbox3 – screenshot SentiCircles '#refugeeswelcome' tweets

The one concern of interviewees was that the lists for core and related terms include many words 'without a meaning' if they appear separately. These could be filtered out or grouped together to get combinations of words that provide greater meaning. While scrolling the list of core terms the MP noted that there were many terms indicating links to soccer: this demonstrated the improved accuracy of the search and analysis as there is substantial mobilisation of soccer fans around the '#refugeeswelcome' message in Germany.

A demonstration case was also built around the NRW acronym (for North Rhine Westphalia). In total 500 tweets were collected from 376 unique users over a four-hour period into the early evening. A search was conducted on the terms "NRW" and "#NRW": the former proved problematic, possibly linked to the timeframe for collection, in that a lot of the content was pornographic material, while the latter linked to lots of sub-terms used on twitter (e.g. refugees). The sentiment for "NRW" was more positive and the sentiment for "#NRW" more negative but it was not clear from the search what underpinned the diversification in sentiment.

The SentiCircle tool was also tested on the interviewee MP's twitter handle, @tosopiratas resulting in 194 tweets from 31 unique users over an eight-day period. The MP himself noted that the sentiment looked 'bad' overall and was interested in the reasons for this: was it because his own tweets had negative sentiment or because other twitter users were discussing him in a negative way?

Although Toolbox v3 and the SentiCircle tool was a marked improvement on previous prototypes, it was still not sufficiently clear how and why results have been arrived at.

The results table also shows Twitter names, hashtags and regular words in combination with the sentiment and the correlation, but these measurements are not intuitively understandable for average, non-technical users.

The need for more filter options was also noted – for example, to show only proper nouns or Twitter names as both lists (both core and related terms) can rapidly become quite long.

4.2 Supra-national level: European Union

4.2.1 Feedback

4.2.1.1 Social media SentiCircles/ sentiment analysis

From the users perspective it was not sufficiently clear how they should run a Twitter search, and that these then needed to be added to the SentiCircles for further analysis. The less tech-savvy end-users pointed out that some explanatory information is needed in each window of the tool to set out the functionalities of each component and what the user can expect to do / to happen.

However, bearing in mind initial end user concern regarding provenance and trustworthiness, the response of users on this score was positive. The clear indications of the exact time, users and tweets collected all contributed to a sense that the tool was transparent in the information provided and therefore worthy of trust. Nonetheless, more explanatory information may be needed to explain the SentiCircle indicators – eg. what a sore of 0.6 means as opposed to 60% on earlier iterations of the toolbox.

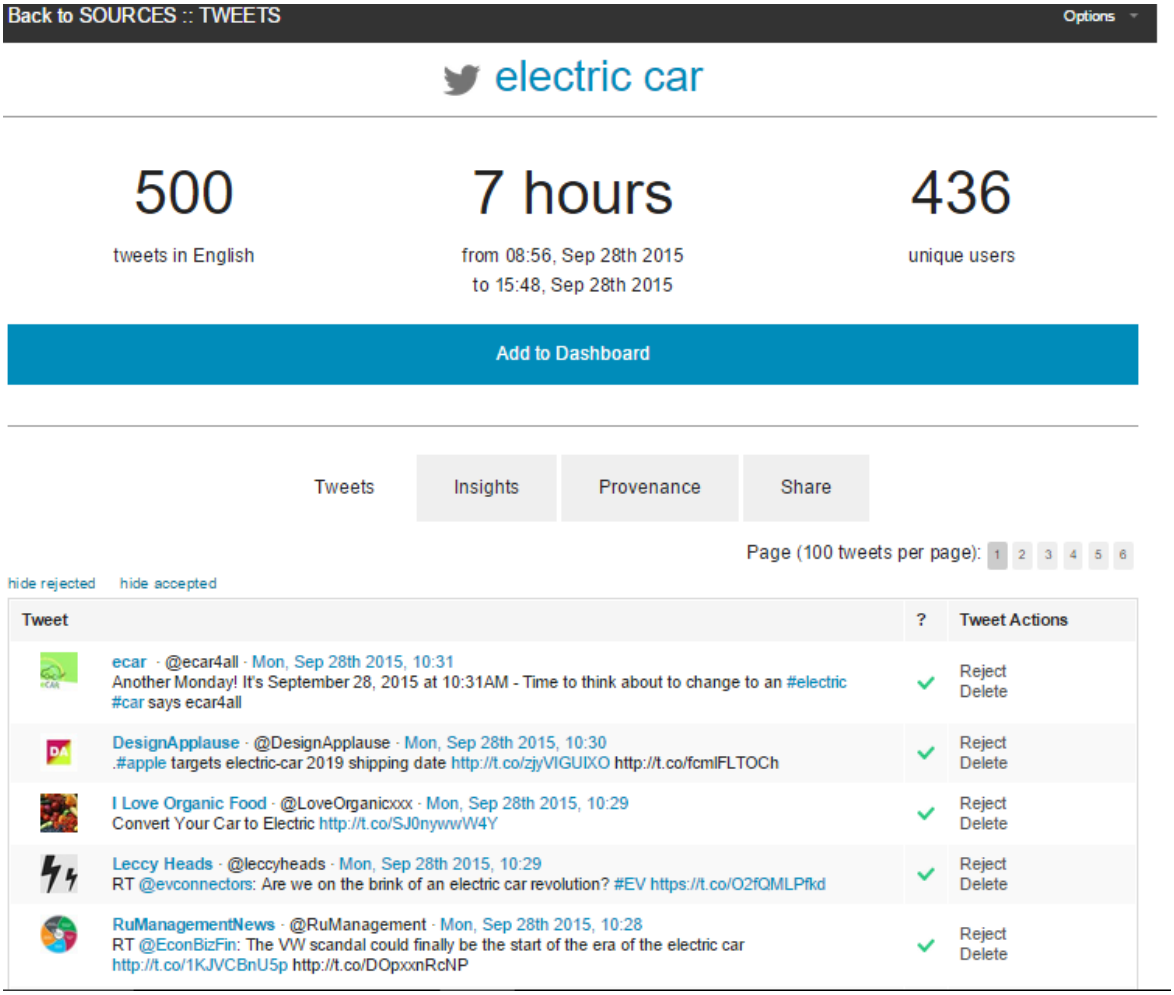

Figure 12: Toolbox3 – screenshot `Electric cars" tweets

One end user interviewee was more familiar than the others with online sentiment analysis and noted that such tools were available via commercial providers but were usually beyond the reach of the public policy sector due to budget constraints even though they could be beneficial for policy making and in certain policy domains could help support reputation

management. The benefit of such analysis was noted in relation to observing views at certain points of time – for example, if the European Commission was about to release public tenders on the promotion of EU and African relations, it would be interested in advance to learn what people are talking about around such relations, what the key issues discussed are, what the sentiment around them is, how the online debate evolves and what would be the best time, shape or angle of any policy campaign that had to be marketed externally.

The other end users were less familiar with the familiar with the sentiment analysis tools and therefore had a fresh reaction to the concept and benefits. They felt it was a very interesting associative disposition of linguistics, that could help generate a good understanding of the main "noise" around a topic on social media.

However, they wanted to be able to run searches at different points in time and to be able to compare these in order to ascertain whether sentiment was changing. They also wished to see tweets for the entire core term – e.g. when it is a word combination such as 'electric car' or 'renewable energy' as splitting and presenting the tweets only per single word, loses the actual sense of the original core term.

Interviewees also reiterated concerns that the tool needs to be adjusted in order to enable the user to identify tweets from unique users and those that are commercial.

Location

All the end-user interviewees agreed that a very important requirement for social media analysis is the *location* choice, while performing a social media search. That means, end-users want to be able to choose location to narrow down the geographical scope of the audience and to be able to identify who stands behind certain tweets. The option to run searches on the same topic, but for different geographical locations, would be a further benefit as it would enable end-users to compare competitive markets. When we talk about policy making on the European scale, it is about member states positions, international competition and positioning, it is therefore vital to be able to categorize the "tweeters". Such comparison opportunity would be a huge analytical advantage, as it would be possible using one tool to actually grasp different member states opinions, to track industry positions and make certain assumptions of a public opinion.

Language

Following the indicated *location* requirement by the end users it also poses an additional *language* requirement for Sense4us prototype. Any information tool that aims to be beneficial for European policy making, must be able to search in local languages.

Inclusion of media analysis?

After getting acquainted with the main components of Toolbox v3, two of the three end users asked whether general media could also be tracked. Policy making is not a fluid process, it is reactive and the traditional media play a significant role in shaping policy makers behaviour and opinions. Therefore, a useful component would be the integration of a timely media search function to provide users with a rapid assessment of sentiment in the mainstream media.

4.2.2 LOD search and LOD surrounding

Due to difficulties posed by the amount of time it was taking to search and load results, the Linked open data search function was not presented in real time with live data – a pre-developed example for presentational purposes was instead. LOD search was entirely new to the end users but they were particularly interested in the concept of 'bridged' data sets.

Generating a map of inter-linked policy issue related data was particularly attractive for research purposes.

However, when the component tools were demonstrated end users were sceptical about what they saw.

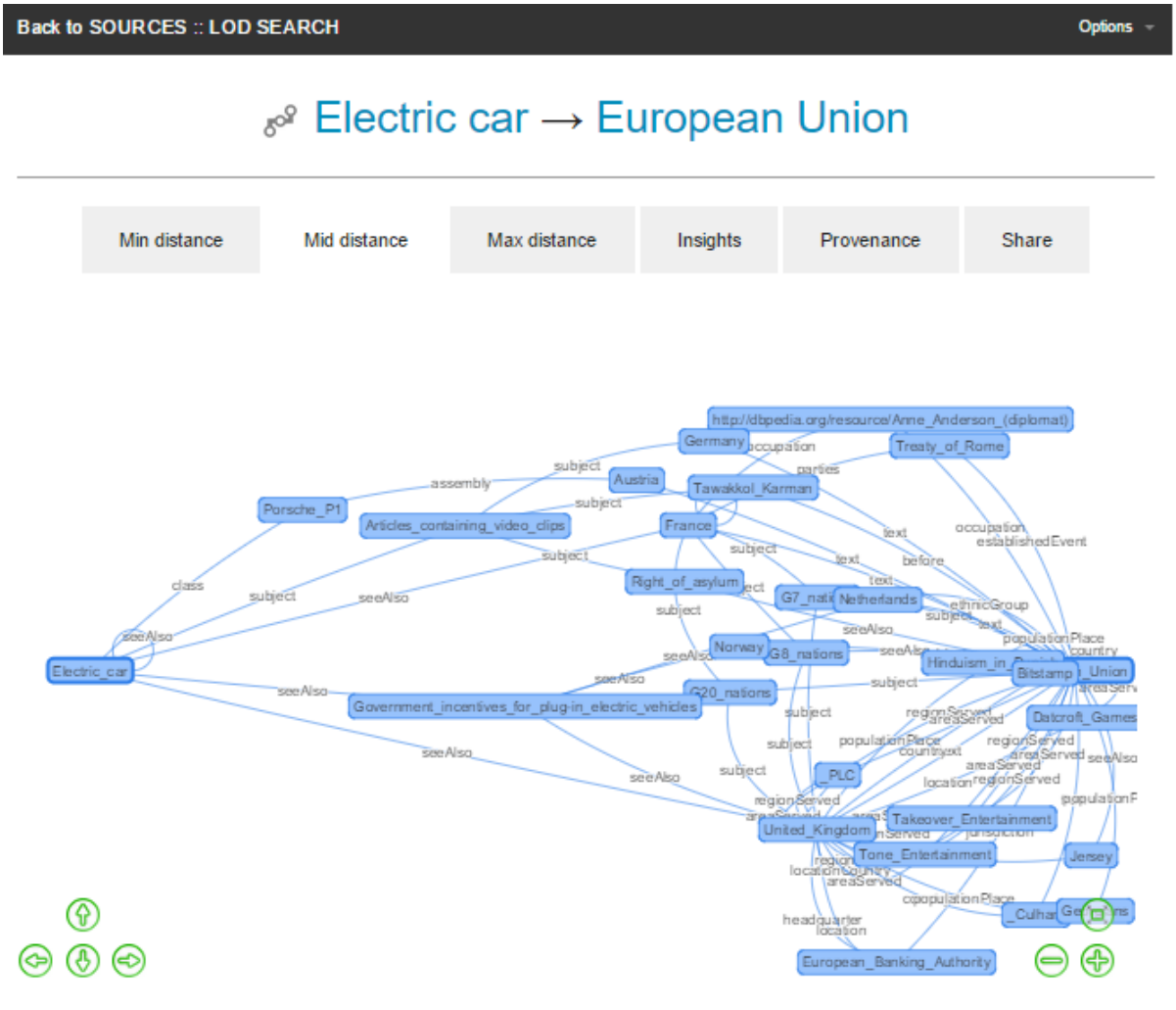

Figure 13: Toolbox3 – screenshot `Electric cars" LOD search

As Figure 13 illustrates, the LOD map for the topic "electric cars", interlinked data such as "right of asylum" or "Hinduism in Punjab", neither of which have readily evident links to electric cars leading the users question the functionality of the component. The search capacity thus requires considerable refinement and the filtering options need improvement.

It was also suggested that a broader range of source datasets be utilised. All end users became critical and sceptical once they learned that the tool's search function draws on Wikipedia. The credibility of information sources is key to policy makers and crowd source editing of publicly available information does not have the analytical value and credibility that many policy makers are looking for. It was suggested that the LOD component be connected to other trusted source databases – e.g. of EU institutions – to draw on different data.

LOD surrounding was welcomed more positively by end users, not least because it provides them with the opportunity to filter information as they see fit. The "map of surrounding" concepts was welcome, and the "table of concepts" (see Figure 14)was deemed useful as a way to enable users to browse through categories and to learn about the data sets in a more structured way.

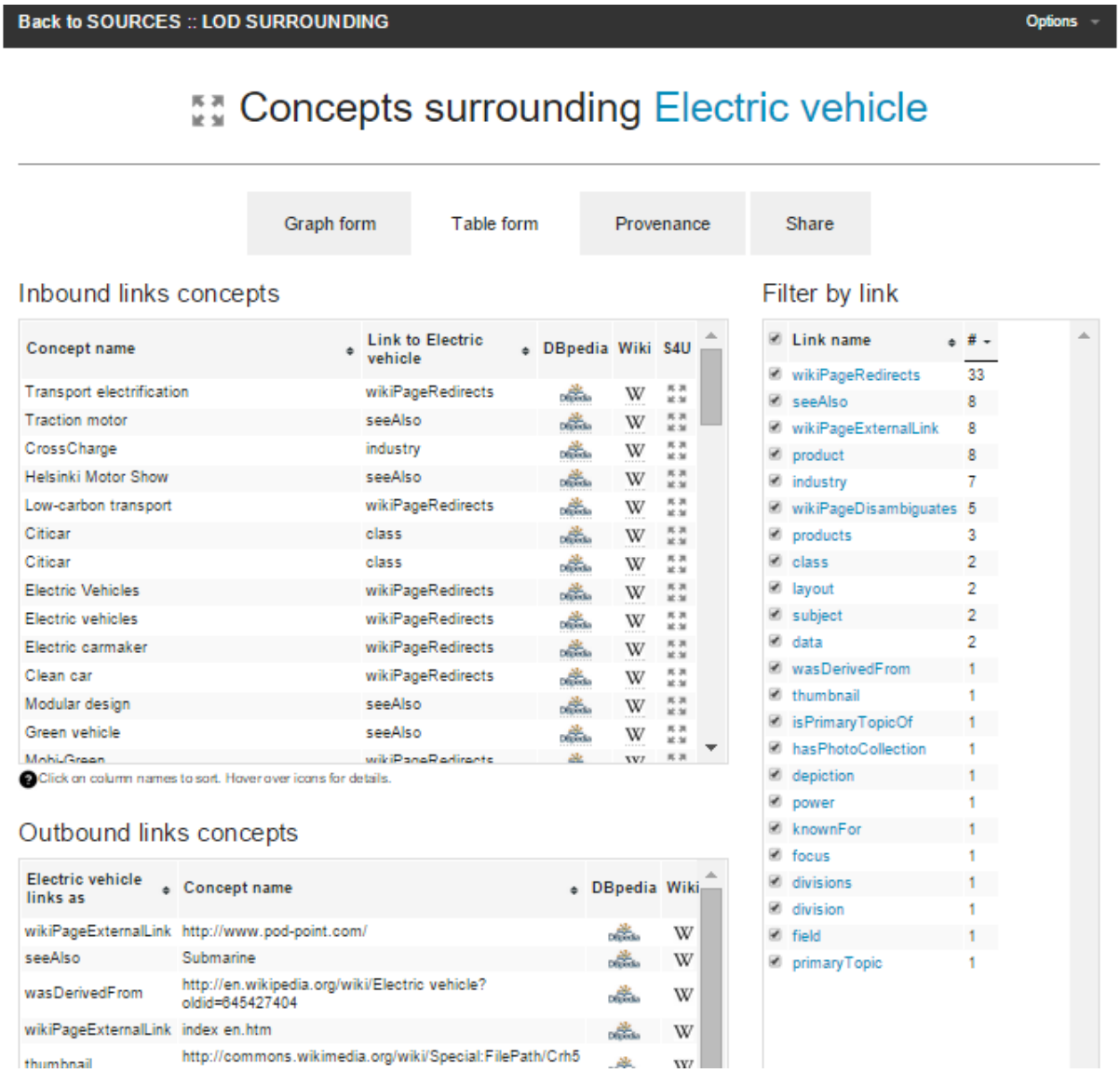

Figure 14: Toolbox3 – screenshot `Electric vehicle" LOD surrounding

4.2.3 Theme overlap

Here EU policy makers were interested in the concept of rapidly analysing a document, being able to quickly identify the main topics discussed, and compare it with social media analysis data. However, the interviewees wanted improvements in the visualisation of the information. And the fact that the themes in the document match the discussions on social media does not automatically general real value – users still have to read the document, or at least key parts of it, in order to understand the context.

5. Case Studies: Potential Deployment Scenarios of Sense4us

In addition to exploring the architecture, functionality and interface of the toolbox, the Hansard Society worked with Edward Faulkener, during his dissolution placement, to develop ideas about how the toolbox as a whole, and individual components on their own, might be utilised in real world scenarios in Parliament and government. Below are two selected case study examples that illustrate the way in which the toolbox could conceivably be used in the future to help the users.

The first is based around the work of a parliamentary select committee, conducting an inquiry into voter engagement. This is modelled in large part on the experience of the House of Commons Political and Constitutional Affairs Select Committee inquiry into voter engagement held in 2014-15.

The second is based on an expansion of the electric cars scenario the end user partners developed in the first year of the project. It looks at the research questions that might be usefully addressed by one or more of the Sense4us search tools.

It is our intention to repeat these test cases on future iterations of the toolbox in order to test development progress.

5.1 Case study 1: Parliamentary committees

Parliamentary committees in the UK are supported by relatively small teams which are responsible for providing briefings and drafting reports, many of which will be on technical subjects of which the staff may have little experience. Tools that enable Committee staff to find relevant information, access public opinion and gauge the impact of a policy would usefully support their core work.

Some of the ways in which Sense4us could be used by Select Committees are outlined below.

Finding and analysing relevant social media debate

Committees are increasingly active in engaging with the public on inquiries – both via social media and other means (informal meetings, surveys, etc.). The Sense4us toolkit could help by using the social media search to identify public debate relevant to their work/inquiry. This could enable:

- Accessing of public opinion about issues the Committee is considering.
- Finding new information that is being discussed online with regard to the Committee's agenda.

Identifying relevant information and organisations

Committees will often be working with a policy document, and will also have access to other information held by Parliament, but in order to draft briefing papers and reports they need to identify other salient material. The Sense4us toolkit could help by:

- Using the Linked Open Data search to identify relevant data and organisations.
- The LOD search function is deemed to have potentially high value in tackling a significant problem facing parliamentary committees. Committee clerks who are not necessarily expert in their committee's subject area, have to locate a representative range of potential evidence givers and witnesses to appear before the committee. An identified difficulty is that of how to reach out beyond the 'usual suspects' and find 'unknown' witnesses relevant to the enquiry. The LOD search function has the potential to help address this.

- Using the social media search to identify organisations, individuals and information relevant to an issue.

For example, a social media search for "voter + engagement" returned details of substantive debate about certain issues (eg. Votes at 16, inequality of voter turnout), individuals / organizations (eg. Operation Black Vote, Bite the Ballot Birmingham) and links to relevant material (eg. an article about automatic registration in California).

Analysing written information

Committees make extensive use of written material, in the form of Government documents, articles, and written evidence produced specifically for the Committee. The Sense4us toolkit could help analysis of this information by:

- Using the theme analysis tool to identify common themes in written evidence and other documents. These themes could then be used identify key topics, and to search for relevant social media comments.

Scrutinising Government policy

The main role of Committees is to scrutinise Government policy. The Sense4us toolkit could aid this by:

- Using the social media search to identify online discussion about Government policy/announcements.
- Using the policy modelling tool to assess the Government's chosen policy option and compare it with alternatives.

Assessing the impact of Committee work

Committees are keen to understand the impact of their work, something which is largely assessed in relation to whether the Government acts on the Committee's conclusions. The Sense4us toolkit could allow for Committee's to consider their impact in other ways by:

- Using the social media search to follow the online response to a Committee's inquiry, evidence or report. This could either be following the Committee actively promoting social media debate, or just by following debate that has occurred naturally.

5.2 Case study 2: Ultra Low Emission Vehicles

Policy aim - the UK Government wants to increase take up of Ultra Low Emission Vehicles (ULEVs) throughout the UK, as part of its wider plans for reducing greenhouse gas emissions.

The Climate Change Act established a legally binding target to reduce the UK's greenhouse gas emissions to at least 80% below base year (1990) levels by 2050. The Act introduced a system of carbon budgets which provide legally binding limits on the amount of emissions that may be produced in successive five-year periods, beginning in 2008. The first three carbon budgets were set in law in May 2009 and require emissions to be reduced by at least 34% below base year levels in 2020. The fourth carbon budget, covering the period 2023–27, was set in law in June 2011 and requires emissions to be reduced by 50% below 1990 levels. Domestic transport emissions make up nearly a quarter of UK emissions (24% in 2009). The Government has stated that, by 2050, domestic transport will need to substantially reduce its emissions.

Part of the UK Government's "vision" for reducing emissions is ultra-low emission vehicles (ULEVs) including fully electric, plug-in hybrid, and fuel cell powered cars. Its report on delivering a low carbon future states:

Over the next decade, average emissions of new cars are set to fall by around a third, primarily through more efficient combustion engines. Sustainable biofuels will also deliver substantial emissions reductions. As deeper cuts are required, vehicles will run on ultra-low emission technologies such as electric batteries, hydrogen fuel cells and plug-in hybrid technology. These vehicles could also help to deliver wider environmental benefits, including improved local air quality and reduced traffic noise.

The Government's policies in this area include:

- pressing for strong EU vehicle emissions standards for 2020 and beyond in order to deliver improvements in conventional vehicle efficiency and give certainty about future markets for ultra-low emission vehicles.
- providing around £300 million in the 2010-15 Parliament for consumer incentives, worth up to £5,000 per car, and further support for the research, development and demonstration of new technologies.

  providing a £560 million Local Sustainable Transport Fund over the lifetime of the 2010-15 Parliament, to support people to make lower carbon travel choices, such as walking, cycling or public transport.

Questions the policy-maker might ask:

1. What discussions are happening online about ULEVs?
    a. What types of ULEV are most popular in social media discussions?
2. What are the perceived barriers to the take up of ULEVs?
    a. What are the most important barriers to the take up of ULEVs?
    b. What are the incentives that might increase the take up of ULEVs?
3. What might effective policy options and messages be?

How Sense4us might help answer these questions:

Considering online discussions about ULEVs

Policy-makers need to know more about public opinion regarding ULEVs, to provide context for the development of policy in this area.

There are several different types of ULEV; fully electric, plug-in hybrid, fuel cell powered cars, and increasingly efficient internal combustion engines. Public sentiment towards certain types of ULEV may vary.

Sense4us could help to answer these questions in a number of different ways:

- The analysis of a policy document might help to scope keywords related to ULEVs which might be used in social media discussions.
- Volume of social media results and sentiment analysis of social media could allow policy-makers to identify which type of ULEV was most popular in online discussions.

Identifying perceived barriers to take up of ULEVs

Public behaviour can be hard to change. Before policy-makers can understand what mechanisms might be effective in increasing the take up of ULEVs, they need to understand the perceived barriers to take up of ULEVs, which groups of people might be willing to make this change and what messages or incentives would be most likely to influence their behaviour.

Sense4us could help to answer these questions in a number of different ways:

- Sentiment analysis from social media could split discussions around ULEVs into themes allowing policy-makers to what issues online discussion centres around.

- Policy-makers could segment groups based on location so that if, for example, there was a particularly strong discussion about ULEVs or a related term in a certain area, that area could be targeted for a pilot or focused messaging.
- Linked open data search of relevant themes and keywords could identify individuals and organisations interested in ULEVs.

Considering policy options

There are a variety of different policy options which policy-formers could use to encourage take up of ULEVs. These include grants for those purchasing ULEVs, emission standards for new cars and additional benefits for ULEV users (eg. Free parking, exemption from tolls, etc). Sense4us could enable policy-formers to decide which policy options to use by:

- Theme analysis of social media comments to identify the policy options which are most discussed online by members of the public, and the views towards these options.
- Creating a policy model to assess how different policy options affected take up of ULEVs, and what impact this would have on greenhouse gas emissions.

Process for testing Sense4us tools:

1. **What discussions are happening online about ULEVs?**

- scoping exercise, identify main themes using policy documents
    - upload "strategy for ultra low emission vehicles in the UK.doc" (UK Government policy document) (Note toolkit v2 only extracted 100 paragraphs of the document so the majority was not extracted)
        - run topic analysis, get main themes:
            - vehicles, chargepoints, infrastructure, charging
            - ulevs, sector, government, energy
            - technology, vehicles, market, low
            - vehicles, fuel, power, engine
    - upload “motoring of the future.doc” (House of Commons Transport Select Committee report)
        - run topic analysis, get main themes:
            - road, government, emissions, set
            - safety, vehicles, road, professor
            - technologies, vehicles, new, drivers
            - dft, data, different, note
- search Twitter using policy document themes
    - for a couple of themes search twitter (note: key words may not be terms that are used on Twitter)
        - Repeated tweets/retweets give less variety in viewpoints
        - Significant presence of commercial/advertisement tweets that need to be deleted
        - Lots of Tweets are not relevant – users can’t just search for Tweets including any of main themes. Need to limit Tweets to those including at least two keywords (or include main subject – ULEVs)

- Many Tweets are clearly not UK-based. But there are too few tweets for analysis if the search is constrained to UK
- A degree of trial and error is necessary to return a reasonable volume of Tweets which provide relevant and useful information

Useful results:

Searching social media for "electric cars" + chargepoints identified discussion around new chargepoints being installed in Worcestershire – including links to an article about infrastructure development resulting from government-funding.

***Value**: identified locality of interest to policy-formers, example of policy impact being responded to on social media.*

Searching social media for "electric cars" + emissions identified debate about the impact of electric cars on greenhouse emissions (with arguments for and against electric cars).

***Value**: identified contrasting public opinion regarding electric cars, helpful in identifying challenges to presenting electric cars to the public (need to respond to arguments).*

- Theme analysis on tweets:
  - Main themes for results from "low emission" + vehicles:
    - telegraph, scenic, summer, roadtrip
    - vehicles, emission, ultra, hoyer
    - emission, low, cost, natwestbusiness
    - emission, zones, polluting, national
  - Doesn't give any obvious insight. Would need to dig into tweets to understand most of the themes
- Sentiment analysis of Tweets
  - "low emission" + vehicles returned 18 positive and 20 negative Tweets. Not clear sentiment assigned by tool reflects common sense reading of Tweet or intent of individual posting. Some "negative" Tweets are clearly positive – eg:
    - Low-emission vehicles: How the cost of going #green may be a lot less than you think...
    - It's quick and easy to arrange a test drive for any of our ultra low emission vehicles, just visit https://t.co/Z3MaeXQO3g
    - Ultra-low emission vehicles: more choice than ever http://t.co/z8hPb8IV9V #News, vehicles
    - Record levels of Ultra Low Emission Vehicles were registered in the first quarter of the year (via @BusinessCar) http://t.co/e5J4KcGknO

What types of ULEV are most popular in social media discussion?

- Social media search for different types of ULEV (eg. Vehicle, car + electric/hybrid/hydrogen)
  - Time period over which 250 Tweets have been posted indicates volume of Tweets relating to particular type of ULEV (eg. 6 hours for "electric", a day for "hybrid" or "hydrogen")

- Theme analysis of Tweets returned for different types of ULEV:

- Mostly picks up on different manufacturers producing certain type of ULEV (relates to point about significant volume of Tweets being commercial/advertisements)

- Sentiment analysis of Tweets by type of ULEV
  - Majority of Tweets (over 200) were positive for each type of ULEV
    - Eyeballing Tweets throws up disparity between actual sentiment and how the tool categorizes them (as with previous use of sentiment analysis)

2. What are the perceived barriers to the take up of ULEVs?

- sentiment analysis from Twitter

  - What are the most important barriers to the take up of ULEVs?
    - Run topic analysis for negative Tweets – no relevant information in "ULEV" Tweets, but negative Tweets returned for "electric cars" include references to relevant barriers/concerns:
      - Charging
      - Expense
      - Negative impact regarding greenhouse gases
        - "afford" and "charge" were returned in a theme analysis of negative Tweets
    - can run new twitter search that includes some of the barriers identified above to get more results/detail
  - What are the incentives that might increase the take up of ULEVs?
    - Topic analysis for positive Tweets doesn't return any significant themes

Useful results:

Analysis of Tweets returned for "electric cars" turned up a number of perceived barriers to take up of electric cars, in addition to debate relevant to the wider question of how ULEVs relate to the issue of greenhouse gas emissions.

3. What might effective policy options be?

- The perceived barriers to take up of ULEVs can be used to identify possible policy options, eg.
  - Charging as a barrier: Look at charging infrastructure
  - Expense as a barrier: Look at how costs can be reduced (grant for purchase of ULEV, free charging, exemption from tolls/parking)
  - Negative impact regarding greenhouse gases: Look at environmental impact of ULEVs and electricity production
- Policy model tool would be useful once possible options are identified

Overall Analysis:

The returns from theme analysis on toolbox v2 were of varying value. Some of them were entirely sound, but may not add significantly to the knowledge of the user, while other

themes did not link keywords in a useful way. A different way of representing key words from documents might be more helpful.

The tool is useful for finding and sorting social media content relevant to policy-makers. This content gives insight into how the public is engaging with issues and policies, helps policy-formers identified interested individuals and organizations and highlights some of the perceived barriers to take up of ULEVs. The sentiment analysis was not terribly accurate (this was improved for toolbox v3) but would be a useful tool if it could be relied upon as it would enable policy-formers to look at commonalities across positive or negative comments.

6. Conclusions

This deliverable has concentrated on feedback from the end users regarding these prototypes and the resulting requirements arising from this feedback. In addition, this deliverable has indicated which tools and which features were seen as useful, and which were not.

6.1 Summary of Overall Key Findings

This section contains some of the key findings and feedback on the tools themselves. The overall reactions for each tool are as follows.

Key *Themes*

The concept of document summarisation is useful and could be very valuable, but the results of this tool are very difficult to understand, as they are collections of words.

Theme Overlap

This tool could be useful, but to be useful needs clearer presentation. For example, what constitutes a paragraph?

Sentiment Analysis

There was a strong consensus that the new "SentiCircles" tool in V3 of the toolbox is a significant improvement over the previous sentiment analysis. The results are more believable and the presentation is clearer.

Twitter Search

Searching social media is clearly useful, as it enables the user to see what people are talking about.

General

The general concept of the toolbox is useful, but the overall workflow is difficult to understand – it is not clear "what to do next".

Policy Modelling and Simulation

The concept of policy modelling was not demonstrated, as it is just becoming ready to show end users at the time of writing, but feedback on the concept itself indicates that the users need clear examples to illustrate:

- The benefits of policy modelling
- How models are built
- How the other tools can contribute to policy modelling
- How the models can be trusted.

6.2 Summary of Key Feedback

Below is a list of the key feeedback that has emerged from the evaluation of Toolboxes 1-3 undertaken so far (testing of version 3 is not yet complete) which need to be considered and addressed. This is not an exhaustive list of all suggestions for new or modified requirements as many minor changes have been agreed and implemented on an iterative basis as a consequence of pre-evaluation testing by end user partners.

Finding: new or modified function requirement

Include feedback / error messages to signpost users
Website script notification alerts
Incorporate add/delete function buttons for all tools throughout
Review use of icons and labelling for all components
Consider an advanced settings mode
Search to extract text from tables in documents
User direction of automatic acceptance/rejection of paragraphs of text
Include option to identify Twitter actors in social media/sentiment analysis of results
Show the 'workings' by which sentiment results have been arrived at
Expand use of online data search beyond Twitter – e.g. to Facebook, blogs etc
Search tweets using 'at least two of these words'
Future tracking/monitoring of social media searches & past history
Location identification of tweets
Filtering out of commercial/inappropriate tweets
Extend sentiment function to documents
Word usage statistics to replace main themes analysis
Drag and drop facility for LOD result maps
Introduce abbreviation / colour coding of links on LOD maps for brevity and incorporate hyperlinks
Expand LOD database search capacity beyond Wikipedia
Non-English language function

Table 5: Summary of end-user feedback & requirements

This feedback will be used to determine the functional requirements for upcoming versions of the toolkit. The functional requirements deriving from the feedback in this deliverable will be reported in D3.3.