

# DIGIT.B4 – Big Data PoC

# DIGIT 01 – Social media topics

D04.02.User manual and results report

everis Spain S.L.U



### Table of contents

| 1 | Sup   | port Information and Points of Contact | 4  |
|---|-------|--|----|
| 2 | Intro | oduction                               | 5  |
|   | 2.1   | Context                                | 5  |
|   | 2.2   | Objective                              | 5  |
| 3 | Syst  | tem Overview                           | 6  |
| 4 | Gett  | ing Started                            | 7  |
| 5 | Usin  | ng the System                          | 9  |
|   | 5.1   | DATA SECTION                           | 9  |
|   | 5.2   | ANALYSIS SECTION1                      | 3  |
|   | 5.2.  | .1 Overview                            | 14 |
|   | 5.2.  | .2 Temporal evolution                  | 20 |
|   | 5.2.  | .3 WordCloud                           | 24 |
|   | 5.2.  | .4 Top terms                           | 27 |
|   | 5.2.  | .5 Excel button                        | 32 |
|   | 5.3   | ABOUT SECTION                          | 3  |
| 6 | Syst  | tem Summary3                           | 5  |
| 7 | Sum   | nmary of Analysis Results3             | 6  |



### Table of figures

| Figure 1- Web tool scheme   | 6                    |
|---|----------------------|
| Figure 2 - Home website   | 7                    |
| Figure 3 - Clicking on Data section   | 8                    |
| Figure 4 - Treemap Section  | 9                    |
| Figure 5 - Treemap section example: Posts categorized as "Digital Transformation"   | 10                   |
| Figure 6 - "Show all posts" button example  | 11                   |
| Figure 7 – Clicking on the title link of the post example   | 12                   |
| Figure 8 – "Next" button example: Posts "Digital Transformation" in second page   | 13                   |
| Figure 9 - Sub-navigation bar   | 14                   |
| Figure 10 - Overview section  | 15                   |
| Figure 11 - Overview section with filters   | 16                   |
| Figure 12 - Overview example: Data of "Big Data" and "Collaborative work" catego  | ories<br>17          |
| Figure 13 - Overview example: Name of a category is too long  | 18                   |
| Figure 14 - Overview example: Amount of posts   | 19                   |
| Figure 15 - Overview example: Download excel  | 20                   |
| Figure 16 - Temporal Evolution section  | 21                   |
| Figure 17 - Temporal Evolution section with filters   | 22                   |
| Figure 18 - Temporal Evolution example: Data of "Big Data", "Collaborative we<br>"Digital transformation", "Electronic Government" and "The cloud in the EU" catego | ork",<br>ories<br>23 |
| Figure 19 - Temporal evolution example: Download excel  | 24                   |
| Figure 20 - Word cloud section  | 25                   |
| Figure 21 - Word cloud section with filters   | 26                   |
| Figure 22 – Word cloud example: Download excel  | 27                   |
| Figure 23 - Top Terms section   | 28                   |
| Figure 24 - Top Terms section with filters  | 29                   |
| Figure 25 - Top terms example: Data of "Collaborative work" categories  | 30                   |
| Figure 26 - Top terms example: Data of "Collaborative work" categories with tooltip   | .31                  |
| Figure 27 - Top terms example: Download excel   | 32                   |
| Figure 28 - Download excel  | 32                   |
| Figure 29 - Example of download excel   | 33                   |
| Figure 30 - About section   | 34                   |
|   |                      |



### 1 SUPPORT INFORMATION AND POINTS OF CONTACT

| Name          | Company  | Email | Telephone number |
|---------------|----------|-------|------------------|
| Marco Fichera | DIGIT.B4 |       |                  |
|               |          |       |                  |



### **2 INTRODUCTION**

### 2.1 Context

This document will specify the purpose of each section of the tool, how to navigate it and how it works.

First of all we will describe the "home" section, followed by the "data" and the "analysis" sections. To conclude, the "about" section will be explained.

\_\_\_\_\_

### 2.2 Objective

I

١

The aim of this document is to describe how all sections of the tool work and the actions that a user can perform on each of them.



### **3 SYSTEM OVERVIEW**

The purpose of the tool is to display different data sets coming from Yammer posts and tweets previously analyzed and categorized in order to facilitate their visualization.

The results of the analysis has been published in a web tool with a standard content distribution.

| HEADER  |         |  |  |  |
|---------|---------|--|--|--|
| FILTERS | CONTENT |  |  |  |
| FOOTER  |         |  |  |  |

**Figure 1- Web tool scheme** 

The user will be able to visualize the data in four different views: Overview, Temporal Evolution, WordCloud and Top Terms. Each one displays different data sets, like the number of posts classified in each category, its temporal evolution per week or the most common terms for a given category.



### 4 GETTING STARTED

Once the user accesses to the tool via the given URL, the home section will be displayed.

The Home section is the entry point to all the sections of the tool. The home page has a navigation bar at the header, which can be used to navigate through the different sections of the tool. The navigation bar is present in all the sections of the tool. These sections are:

- Home
- Data
- Analysis
- About



#### Social Media

This proof of concept (PoC) showcases the usefulness of the study and analysis of the social media domain. This PoC demonstrates the use of text mining techniques on social media as a means to identify areas of interest in research, specifically to gather conclusions regarding the most important topics to discuss during the ICT conference 2016. The information to analyse has been gathered from Twitter and Yammer posts concerning previous events.



Figure 2 - Home website

There is an introductory paragraph under the navigation bar, describing the reasons why this tool has been developed and the objectives that it tried to accomplish.

There is an alternative way to navigate through the tool in the center of the home section. This alternative option has three different icons that correspond to the different sections (Data, Analysis and About) of the tool.

The user can access each section by clicking on the icons or on the relative section name in the navigation bar.



Figure 3 - Clicking on Data section

At the footer of the home page different links can be found. The links take the user to additional informative sections of the tool, as well as to the website of the entities involved in this project. Just like the header with the navigation bar, the footer is present in all the tool's pages. The links in the footer are:

- Policies and Guidelines
- Contact
- European Commission
- Everis



### 5 USING THE SYSTEM

### 5.1 DATA SECTION

The data section is where the tool lists all the Yammer posts and Tweets involved in the analysis. The purpose of this section is to let the user see the posts and select and filter them by category.

Once the user accesses this section, he will be able to see two different parts:

- The Treemap with the categories (left side)
- The list of selected items (right side)



**Figure 4 - Treemap Section** 

The treemap shows the categories defined for the classification of the posts. Categories with lighter color have less associated posts than the ones with darker colors, and their size is smaller.



The user can filter the list of posts by clicking on a category of the treemap. Once the user has clicked on it the posts belonging to this category will be shown on the left side. Only one category can be picked at the same time.

The figure below shows the results once the user has clicked on the category "Digital transformation"

| DIGIT - So              | DIGIT - Social Media PoC                         |                      |  |  |  |  |
|-------------------------|--|----------------------|--|--|--|--|
| 🟫 НОМЕ                  |  | 🛷 data               | ANALYSIS     ABOUT   |  |  |  |
| other                   | Show all posts<br>Digital tran<br>Digital transf | nsformation          | Z75 results found for Digital transformation       Best digital innovation award for #digitconf, w DenisProst1 @webmutation @TweeTikitin #internationms @EU_Commission       Deni Simoneer 1921 - 18 inn 2015       Category: digital_transformation       Com poden lea #TIC Impulsari la transformació digital al sector public? http://ow.ly/EYFbN #DiGiToonf @EU_Commission       Consorei AOC - 11:45 pm - 10 dec. 2014       Category: digital_transformation       "If the public sector len't digital, Europe len't digital" says @GOettingerEU to @EU_Commission staff at #DiGiToonf via @DiGiToonf       "arco Recorder - 5:00 am - 19 nov. 2014 |  |  |  |
|                         | Discussion and learning                          |                      | <u>Category</u> : digital_transformation<br>Out of all conferences on #digital of past 2 yrs in BRU, #DIGITconf was by far the best, the most<br>profeesional and the least promotionall<br><i>Kristina Jasaity</i> te - 8:49 am - 18 nov. 2014  |  |  |  |
| New communication tools | ommunication tools Culture of change             |                      | <u>Category</u> : digital_transformation<br>My #DIGITConf final tweet: good experience, useful day & nice experience 'accompanying an event'<br>through tweets visible on the ecreen.<br><i>Martin Bohle</i> - 8:40 am - 18 nov. 2014<br><u>Category</u> : digital_transformation<br>#DIGITConf: digital = rethink business<br><i>Martin Bohle</i> - 8:09 am - 18 nov. 2014<br>Category digital_transformation   |  |  |  |
| People engagement       | Electronic Government                            |                      | How to get citizens more involved into the EU policy-making process? Answers at @H4P2014<br>http://oww.lyiEtdos #DiGiTconf #HÆUP<br>Digital Agenda - 8:08 am - 18 nov. 2014<br><u>Caregory</u> : digital_transformation  |  |  |  |
| Big Data                | Innovation and se                                | curity Collaborative | "We have to tap into the power of the millennials - we need reverse mentoring in @EU_Commission"<br>Director-General @stephen_quest #DiGiTconf<br>DiGiTconference - 8:05 am - 18 nov. 2014<br><u>Category:</u> digital_transformation<br>My take: Digital is default, not up for discussion. People and users are where our focus should be.<br>Connect digital and analog! #DiGiTconf   |  |  |  |
|                         |  |                      | Page 1 of 31 next  |  |  |  |
|                         |  |                      |  |  |  |  |

Figure 5 - Treemap section example: Posts categorized as "Digital Transformation"

The user can return to the initial list of posts by clicking on "show all posts" button.



| A HOME                           |                                       | 🔊 Data        | ANALYSIS ABOUT   |
|----------------------------------|---------------------------------------|---------------|--|
|                                  |                                       |               |  |
| Show all posts<br>Show all posts |                                       |               | 1926 results found for all categories  |
|                                  | Digital tran                          | nsformation   | Best digital Innovation award for #digitconf, w DenisProst1 @webmutation @TweeTikitin<br>#Internalcomms @EU_Commission<br>Deni Simeonova - 12:21 - 16 jun. 2015<br><u>Category</u> : digital_transformation  |
| Other                            |                                       |               | Plioting reverse mentoring by millennials @EU_Commission. Thanks to @mjmorrow for the<br>inspirationi #DiGitroonf<br>Stephen Quest- 4:43 am - 5 jun. 2015<br><u>Caregory</u> : new_communication_tools   |
|                                  | Discussion and learning               |               | #digitconf 2014, #walkthetalk, it's a long way, but DIGIT starts walking. Well done, keep walking!<br>Roberto Sgnaolin - 11:02 pm - 18 may. 2015<br><u>Caregory</u> : people_engagement<br>@moraddock to be fair @MTBracken was @EU_Commission #digitconf last Nov presenting #GaaP<br>Mark K Smitham - 1:02 pm - 4 mar. 2015<br>Caregory other  |
| ew communication tools           | Culture of change The cloud in the EU |               | Dear @kodel - please see my previous tweets about @proximus treatment. Hoved your talk at the #DIGITconf in Nov, I hope you can assistituee<br>Ryan Heath - 2:48 am - 17 jan. 2015<br><u>Category</u> : discussion_and_learning<br>Raising Leaders. via @karacegabor: #DIGITconf<br>Sylvia Alexander - 8:28 am - 15 dec. 2014<br><u>Category</u> : culture_of_change<br>Theaks @stephen_guest for the guote bith: //surgos.eu/!//m3/mE_ too much bogget Lealeved |
| People engagement                | Electron                              | ic Government | #digition if a contributing. Looking for actions :)<br>Fabio Zilberstein - 3:15 am - 11 dec. 2014<br><u>Category</u> : other<br>Com poden les #TIC Impulsar la transformació digital al sector públic? http://ow.ly/EYFbN<br>#DIGIToonf @EU_Commission   |
| Big Data                         | Innovation and security Collaborative |               | Consorci AOC - 11:45 pm - 10 dec. 2014<br><u>Caregory</u> : digital_transformation<br>Debats I reflexions sobre la #Innovació a les #AAPP, ICT 2014 - Digital Now! http://ow.byIEYFIb<br>#DIGITconf @EU_Commission<br>Consorci AOC - 3:45 am - 9 dec. 2014<br>Caregory: other  |
|                                  |                                       |               | Page 1 of 214 next   |

Figure 6 - "Show all posts" button example

Each post of the list contains the following specific information:

- **Post**: it is the whole description of the post shown (either Yammer or Twitter). If a user clicks on it, he will be redirected to the original source of the post.
- **Author**: it is the name of the person who has created the post.
- **Date:** it is the date when the post was created.
- **Category:** it is the classification which the item is related to.

If the user clicks on the title link of the post, a new tab will open in the web browser with the related Yammer or Twitter page (as appropriate), thus allowing the user to access the original content.



Figure 7 – Clicking on the title link of the post example

Only nine posts are displayed in each page. It is possible to native through the entire set of posts by using the paginator section under the list of posts:

- "Next" button will display the next nine posts
- "Previous" button will display the previous nine posts



| 🟫 НОМЕ                      |                                       | ~                      | DATA          | ANALYSIS ABOUT   |  |  |
|-----------------------------|---------------------------------------|------------------------|---------------|--|--|--|
|                             | thow all posts                        |                        |               | 275 results found for Digital transformation   |  |  |
| Other                       | Digitai trar                          | Digital transformation |               | #DIGIToon1 It's easy: Just do the right thing (and deliver it, of course)<br>HeinzAchmed - 8:01 am - 18 nov. 2014<br><u>Category</u> : digital_transformation<br>#DIGIToon1 Closing messages: DIGIT is committed to drive the change to Digital by Default and to<br>lead by example<br>Joao Rodrigues Frade - 8:01 am - 18 nov. 2014<br><u>Category</u> : digital_transformation                |  |  |
|                             | Discussion and learning               |                        | ing           | You are at the heart of the Digital Transformation. Comm. Oettinger at #DiGiTconf<br><i>Pierre A. Damas</i> - 8:00 am - 18 nov. 2014<br><u>Caregory</u> : digital_transformation<br>"You don't need physical control of your territory to provide services to your citizens." #Digitconf<br>@juukalives<br>tec.goblin - 7:52 am - 18 nov. 2014<br><u>Caregory</u> : digital_transformation       |  |  |
| New communication tools     | on tools Culture of change            |                        | sud in the EU | A view from the track about transformation, picture not excellent, but @gerrymogovern<br>#DIGITconf<br>Beto - 7:50 am - 18 nov. 2014<br><u>Caregory</u> : digital_transformation<br>Counsellor for Digital Affairs @uukaelives talking about the Estonian digital transformation<br>experience #DIGITconf<br>Kelly Liljerno - 7:42 am - 18 nov. 2014<br><u>Caregory</u> : digital_transformation |  |  |
| People engagement           | Electronic Government                 |                        | ment          | #DIGITconf @uukasilves sharing the Estonian experience of digital transformation<br>Stef Dzhumalieva - 7:38 am - 18 nov. 2014<br><u>Category</u> : digital_transformation<br>Reminders of this day at #digitconf: "engage the users" and we're all "natural born hacker monkeys"<br>Arnaud Cramazou - 7:08 am - 18 nov. 2014   |  |  |
| Big Data                    | Innovation and security Collaborative |                        | ollaborative  | Category: digital_transformation<br>After some 1935 twitts and 178 yammer and growing, do you feel more digital now? Are you<br>becoming a now-lat? #DIGITconf<br>Beto - 7:32 am - 18 nov. 2014<br>Category: digital_transformation  |  |  |
|                             |                                       |                        |               | previous Page 2 of 31 next   |  |  |
| Bollelar and Guidallane 1.0 |                                       |                        |               |  |  |  |



### 5.2 ANALYSIS SECTION

The Analysis section is where the main purpose of the PoC is implemented. It has an additional navigation bar just below the main navigation bar. This secondary navigation bar comprises four subsections that let the user enter into different pages of the analysis section, which are:

- Overview
- Temporal evolution
- WordCloud
- Top terms

|                |           | everis<br>an NTT DATA Company |  |  |
|----------------|-----------|-------------------------------|--|--|
|                |           | T                             |  |  |
| DIGIT - Social | Media PoC |                               |  |  |
| DIGIT - Social | Media PoC | C ANALYSIS                    |  |  |

#### **Figure 9 - Sub-navigation bar**

Each of these subsections contains a different way to interpret the analysis performed over the Tweets and Yammer posts. These interpretations use different types of charts and graphs easier to be understood by non-technical users.

Each of the subsections is divided in 2 parts:

- The left part is a filter panel that lets the user limit and specify the results that are shown on the right section.
- The right section is the visual representation of the categories selected on the left panel.

Each subsection shows the results from a different perspective.

#### 5.2.1 Overview

The Overview visualization shows the amount of posts classified as a given set of categories until a specific date.

If the user click on the tab overview, the results are displayed for all categories.

Every bubble corresponds to a category. The size of the bubbles indicates the amount of posts belonging to each category.



|   | and a second |  |  |
|---|--|--|--|
| A HOME  | 📣 DATA   | 🚯 ANALYSIS   | ABOUT  |
| Overview  | Temporal Evolution   | Word Cloud   | Top Terms  |
|   |  |  |  |
| Filters   | Download Overview data   | 5  |  |
| Select Categories<br>Select the categories from one of the lists  | The Overview visualization shows the amo<br>categories "Network" and "Data Managem                             | ount of posts classified as a given set of cat<br>ent" were selected, the graph would show t | egories until a specific date. For example, if only the<br>he amount of posts belonging to the categories "Netwo |
| <ul> <li>Categories</li> </ul>  | and "Data Management".   |  |  |
| Big Data     Collaborative work     Collaborative work     Collaborative work     Collaborative of change     Digital transformation     Discussion and learning     Electronic Government     Innovation and security     New communication tools     Other     People engagement     Prople engagement     Filter |  | People engagemen   | Culture of change  |
| Showing results for<br>Categories<br>+ Big Data<br>+ Collaborative work<br>+ Culture of change<br>+ Digital transformation<br>+ Discussion and learning<br>+ Electronic Government<br>+ Innovation and security<br>+ New communication tools<br>+ Other<br>+ People engagement<br>+ The cloud in the EU             | Discussion and lea   | Electronic Gover In<br>other   | novation and<br>Collaborativ<br>Digital transformation   |
| Policies and Guidelines   Contact   | <b>1</b>   | ri)  |  |

#### **Figure 10 - Overview section**

On the left panel of this section either one or several categories can be picked to show the results. After clicking on the "Filter" button a text listing the selected categories will appear on the bottom of the left panel.





#### Figure 11 - Overview section with filters

Once the user has filtered the categories by using the left panel, the results will be displayed on the right one.





Figure 12 - Overview example: Data of "Big Data" and "Collaborative work" categories

If the name of a category is too long and it does not fit the size of the bubble, it will be cut off.



| 🔒 НОМЕ  | "≁y° data   | 🕀 ANALYSIS  | ABOUT   |
|---|---|---|---|
| Overview  | Temporal Evolution  | Word Cloud  | Top Terms   |
| Filters   | Download Overview data  | XLS   |   |
| Select Categories<br>Select the categories from one of the lists  | The Overview visualization show<br>the categories "Network" and "Da<br>categories "Network" and "Data | rs the amount of posts classified as a give<br>ata Management" were selected, the grap<br>Management" | n set of categories until a specific date. For example, if o<br>h would show the amount of posts belonging to the |
| <ul> <li>Categories</li> </ul>  | categories retarone and bata  | in a goment .   |   |
| <ul> <li>Big Data</li> <li>Collaborative work</li> <li>Cututure of change</li> <li>Digital transformation</li> <li>Discussion and learning</li> <li>Electronic Government</li> <li>Innovation and security</li> <li>New communication tools</li> <li>Other</li> <li>People engagement</li> <li>The cloud in the EU</li> </ul>   |   | People engag  | gemen<br>Culture of change<br>munication  |
| Showing results for<br>Categories<br>+ Big Data<br>+ Collaborative work<br>+ Collaborative work<br>+ Collaborative work<br>+ Discussion and learning<br>+ Electronic Government<br>+ Electronic Government<br>+ Innovation and security<br>+ New communication tools<br>+ Other<br>+ People engagement<br>+ The cloud in the EU | Discussion  | and learnin<br>Other  | Innovation and<br>Collaborativ<br>Digital transformation  |
| Policies and Guidenes   Contact   |   | everis  |   |

Figure 13 - Overview example: Name of a category is too long

Additionally, a tooltip will appear for each bubble if the user hovers over it. The tooltip shows the full name of the category and the amount of posts belonging to it.





Figure 14 - Overview example: Amount of posts

It is also possible to download the results of the filter by clicking on the "XLS" button which is on the top left side of the right panel (<u>See Section 5.2.5 Excel button</u>).





Figure 15 - Overview example: Download excel

#### 5.2.2 **Temporal evolution**

The temporal evolution section provides a common line chart to show the evolution of posts published under a set of selected categories.

If the user click on the tab Temporal evolution, the results are displayed for all categories.

Every line showed corresponds to a category. Each bubble inside the lines indicates the amount of post belonging to this date.







#### **Figure 16 - Temporal Evolution section**

On the left panel of this section either one or several categories can be picked to show the results. After clicking on the "Filter" button a text listing the picked categories will appear on the bottom of the left panel.



Figure 17 - Temporal Evolution section with filters

Once the user has filtered the categories by using the left panel, the results will be displayed on the right one.



Policies and Guidenes | Contact

Figure 18 - Temporal Evolution example: Data of "Big Data", "Collaborative work", "Digital transformation", "Electronic Government" and "The cloud in the EU" categories

It is also possible to download the results of the filter by clicking on the "XLS" button which is on the top left side of the right panel (<u>See Section 5.2.5 Excel button</u>).







#### Figure 19 - Temporal evolution example: Download excel

#### 5.2.3 WordCloud

The Wordcloud section provides a visualization that shows the most common terms for a given set of categories.

If the user click on the tab WordCloud, the results displayed are the terms of the first category sorted by alphabetical order ("Big Data")

Every word corresponds to a term of the selected category. Word size indicates the number of times the term appears on the posts belonging to the selected category.



| A HOME  | 🔊 DATA   | ANALYSIS                                  | ABOUT                         |  |
|---|--|---|-------------------------------|--|
| Overview  | Temporal Evolution   | Word Cloud                                | Top Terms                     |  |
| Overview  Filters  Select Categories  Getext the categories from one of the lists  Categories  Getext the categories from one of the lists  Categories  Getext the categories from one of the lists  Categories  Filter  Categories  Hig Data  Categories  Hig Data | Temporal Evolution     Word Cloud     Top Terms       Download Word Cloud data     Image: Strategy of the most common terms for a given set of categories. The chart will provide only the terms that supass a specific threshold that refers to the amount of times that this term have been found trough the data. |   |                               |  |
| Policies and Guidelines   Contact   | web  | native mistake year<br>machine grant team | discover <sup>2</sup><br>life |  |

#### Figure 20 - Word cloud section

On the left panel of this section either one or several categories can be picked to show the results. After clicking on the "Filter" button a text listing the picked categories will appear on the bottom of the left panel.



Figure 21 - Word cloud section with filters

Once the user has filtered the categories by using the left panel, the results will be displayed on the right one.

It is also possible to download the results of the filter by clicking on the "XLS" button which is on the top left side of the right panel (<u>See Section 5.2.5 Excel button</u>).





Figure 22 – Word cloud example: Download excel

#### 5.2.4 Top terms

The Top Terms section provides a visualization that shows the top twenty terms for a set of a given categories.

If the user click on the tab Top Term, the results displayed are the top twenty terms of the first category sorted by alphabetical order ("Big Data")

Every bar corresponds to a term of the selected category. The length of the bar indicates the number of times the term appears on the posts belonging to the selected category.



#### **DIGIT - Social Media PoC** A HOME ANALYSIS ABOUT Overview Temporal Evolution Word Cloud Top Terms Filters Download Top Terms data XI.S Select Categories Select the categories from one of the lists The Top Terms will provide the top 20 terms for a given set of categories in a simple bar chart. Categories reps 🕑 Big Data 60 80 Collaborative work Digital transformation Discussion and learning Electronic Government Innovation and security New communication tool People engagement inication tools The cloud in the EU Filter worl US Showing results for risk og) Categories + Blg Data record blg categories **Dense**



#### Figure 23 - Top Terms section

On the left panel of this section either one or several categories can be picked to show the results. After clicking on the "Filter" button a text listing the picked categories will appear on the bottom of the left panel.



Figure 24 - Top Terms section with filters

Once the user has filtered the categories by using the left panel, the results will be displayed on the right one.





Figure 25 - Top terms example: Data of "Collaborative work" categories

Additionally, a tooltip will appear for each line if the user hovers over it. This tooltip shows the value the repeats of a term,







Figure 26 - Top terms example: Data of "Collaborative work" categories with tooltip

It is also possible to download the results of the filter by clicking on the "XLS" button which is on the top left side of the right panel (<u>See Section 5.2.5 Excel button</u>).









#### 5.2.5 Excel button

The button named "XLS" lets the user download in excel format the categories picked with the data related to it.

| DIGIT - Social Media PoC |                        |            |           |
|--------------------------|------------------------|------------|-----------|
| A HOME                   | 📣 Data                 | ANALY SIS  | ABOUT     |
| Overview                 | Temporal Evolution     | Word Cloud | Top Terms |
| ïlters                   | Download Overview data | XLS        |           |

#### Figure 28 - Download excel

The user will be able to use the XLS button to download the data result in XLS format file in each display.



|     | 🚽 🔊 • (° •   =                                    |            |                     |       |      |                        |
|-----|---|------------|---------------------|-------|------|------------------------|
| Arc | hivo Inicio Inser                                 | tar        | Diseño d            | e pág | jina | Fórmu                  |
| Peg | a K Cortar<br>Copiar →<br>gar<br>✓ Copiar formato | Arial<br>N | <u>K</u> <u>s</u> - | •     | 10   | ∗ A* A<br>& ∗ <u>A</u> |
|     | Portapapeles 🛛 🖓                                  |            | Fu                  | ente  |      |                        |
|     | D9 🔫  | 0          | $f_x$               |       |      |                        |
|     | A   |            | В                   |       |      | С                      |
| 1   | Category  |            | Repetition          | 15    |      |                        |
| 2   | new_communication_t                               | 169        |                     |       |      |                        |
| 3   | electronic_government                             | 138        |                     |       |      |                        |
| 4   | collaborative_work                                | 84         |                     |       |      |                        |
| 5   | culture_of_change                                 | 148        |                     |       |      |                        |
| 6   | people_engagement                                 | 139        |                     |       |      |                        |
| 7   | big_data  | 138        |                     |       |      |                        |
| 8   | the_cloud_in_the_EU                               | 140        |                     |       |      |                        |
| 9   | discussion_and_learning                           |            | 252                 |       |      |                        |
| 10  | innovation_and_securi                             | 123        |                     |       |      |                        |
| 11  | other   | 320        |                     |       |      |                        |
| 12  | digital_transformation                            |            | 275                 |       |      |                        |
| 13  |   |            |                     |       |      |                        |
| 14  |   |            |                     |       |      |                        |
| 15  |   |            |                     |       |      |                        |
| 16  |   |            |                     |       |      |                        |
| 17  |   |            |                     |       |      |                        |
| 18  |   |            |                     |       |      |                        |
| 19  |   |            |                     |       |      |                        |
| 20  |   |            |                     |       |      |                        |

Figure 29 - Example of download excel

### 5.3 ABOUT SECTION

The "About" section is a static section whose purpose is to explain, without going into much detail, how the analysis process was done for all the Tweets and Yammer posts. It is divided in 4 paragraphs:

- Introduction
- Data sources
- Methodological approach
- Technical arquitecture



O ABOUT

#### ABOUT

#### INTRODUCTION

The execution of this proof of concept (PoC) showcases the usefulness of the study and analysis of the social media domain. This proof of concept demonstrates the use of text mining techniques on social media as a means to identify areas of interest in research, specifically to gather conclusions regarding the most important topics to discuss during the ICT conference 2016. The information to analyse is gathered from Twitter and Yammer posts concerning previous events.

#### DATA SOURCES

The data for the PoC is obtained directly from Twitter and Yammer. This ensures the usefulness and authenticity of the data for its analysis and classification. Even though different APIs exist for both Twitter and Yammer, their restrictions (Twitter for example does not allow users to get tweets published before a week time lapse) make them not feasible to be used in the context of the PoC. Data is obtained manually. Lastly, additional data extracted from one survey isunched on October 2015 is incorporated. This surveys objective was to gather information related to the topics that the people are more interested in to be part of the next ICT conference 2016 agenda.

#### METHODOLOGICAL APPROACH

- From a methodological point of view, the main tasks that have been carried out are: Texh-ming treatment: transform the abstract of the documents into a format that serves as an input for modeling algorithms Clustering: make homogenous groups of similar documents Classification model obtain the rules to classify a document in one of the categories defined Draw conclusions from the classification of documents
- - Classification Unstructured Text-mining Clustering Concl data treatment model . 🖒

#### TECHNICAL ARCHITECTURE

The reference architecture is a collection of building blocks which consist of a series of technologies and methodologies. NTT DATA and events Bg Data Reference Architecture is composed of 2 levels: 1. Layer: Is a group of building blocks which have the same technology/methodology. There are 8 layers that are categorized for all technologies important for the BIG DATA Architecture. 2. Building Blocks: each one represents a technology/methodology. There are 38 building blocks in the BIG DATA Architecture.

Over the base Bg Data Architecture the blocks that have been used in the PoC for each layer have been highlighted.





#### **Figure 30 - About section**



## 6 SYSTEM SUMMARY

The following table shows the different functions that the web tool provides.

| Functionality                | Description   |
|------------------------------|---|
| Visualize publications       | Visualize all the publications used in the analysis                           |
| Access the posts             | Access the Twitter or Yammer posts used in the analysis                       |
| Filter publications          | Filter the results displayed by topic   |
| Download data results in XLS | Download the data result in a XLS format file in each display                 |
| Visualize Overview data      | Visualize the amount of posts classified as a given set of categories         |
| Visualize temporal evolution | Visualize the evolution of posts published under a set of selected categories |
| Visualize word cloud data    | Visualize the most common terms for a given set of categories in a word cloud |
| Visualize top terms          | Visualize the top 20 terms for a given set of categories in a bar chart       |

 Table 1 - Summary of system functionality



## 7 SUMMARY OF ANALYSIS RESULTS

We can conclude that the results more relevant are shown in the next table where the amount of posts classified for each category and its temporal evolution are analysed. In addition, the most common terms for each category have also been studied.

| Conclusion  | Explanation   |
|---|---|
| Most of the posts do not provide any information.   | About the 16% of the posts are categorized as "other".  |
| The most popular topics have been these "Digital transformation" and "Digital and learning".  | "Digital transformation" and "Digital and learning" categories contains each one the 13% of the total amount of posts.  |
| The less popular topic has been "Collaborate work".   | The category with less posts categorized in has been<br>"Collaborate work".   |
| The higher participation of the users took place on the dates next to the ICT conference      | The temporal evolution graph shows that most of the publications were done between $18^{Th}$ of November and $2^{nd}$ of December.  |
| The peak of the participation took place on the date of the ICT Conference.                   | The temporal evolution graph shows that the highest number of publications were done on the 25 <sup>th</sup> of November.   |
| The top term for "Big data" category is "data".   | The term "data" under the "Big data" category appears<br>two of times the next term repeated the most which is<br>"information".  |
| The top terms for "Collaborative work" category are<br>"work" and "commission".               | The terms "work" and "commission" under the "Collaborative work" category appears two times more than the next term repeated the most which is "information".                       |
| The top terms for "Culture of change" category are<br>"change" and "translation".             | The terms "change" and "translation" under the a<br>"Culture of change" category appears two more times<br>than the next term repeated the most which is<br>"information".          |
| The top terms for "Digital transformation" category are<br>"digital" and "transformation".    | The terms "digital" and "transformation" under the a<br>"Digital transformation" category appear many more<br>times than the next term repeated the most which is<br>"information". |
| The top terms for "Discussing and learning" category are<br>"public", "interest", and "learn" | The terms "public", "interest" and "learn" under the a "Discussion and learning" category appear many more times the next term repeated the most which is "information".            |
| The top term for "Electronic Government" category is<br>"government".                         | The term "government" under the "Electronic Government" category appears two of times the next term repeated the most which is "information".                                       |
| The top term for "Innovation and security" category is<br>"innovation".                       | The term "innovation" under the "Innovation and security" category appears two more times than the next term repeated the most which is "information".                              |
| The top term for "New communication tools" category is "use".                                 | The term "use" under the a "New communication tools" category appears two more times than the next term repeated the most which is "information".                                   |
| The top terms for "People engagement" category are "idea" and "people".                       | The terms "idea" and "people" under the a "People<br>engagement" category appear two more times than the<br>next term repeated the most which is "information".                     |
| The top terms for "The cloud in the EU" category are "cloud" and "European union".            | The terms "cloud" and "European union" under the "The<br>cloud in the EU" category appear two more times than<br>the next term repeated the most which is "information".            |
| The most interesting topic for the users is "Digital<br>transformation"                       | The term "digital" belonging to "Digital Transformation" is the most repeated and appears 239 times.  |