SC352DI07171

D02.04 Report on open data publishing and visualisation

Data visualisations

Document Metadata

Date	2017-04-28
Status	For acceptance
Version	0.02
Authors	Alex Vermeylen – PwC EU Services Eva Cobos Cortina – PwC EU Services Daniel Brulé – PwC EU Services Carmen Vandeloo – PwC EU Services
Reviewed by	Nikolaos Loutas – PwC EU Services
Approved by	

This report was prepared for the ISA2 Programme by:

PwC EU Services

Disclaimer:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

This report has been written in the context of "ISA² Development of an open data service, support and training package in the area of linked open data, data visualisation and persistent identification (Action 2016.18)" and documents the visualisations realised in the context of "Task-02.4: Report on open data publishing and visualisation".

The goal is to provide data publishers from the EU Institutions a service that will enable them to further open up and visualise their data in meaningful ways.

In the context, eight sample visualisations of three high value datasets, namely CORDIS[1], Transparency registered[2] and Tenders Electronic Daily[3]) have been created. The data visualisations were built to look at the selected datasets from different perspectives. Each visualisation is supported by a short text in the form of a data story which helps communicate its message to it, target audience.

The selection of datasets has been done according to the following criteria:

- Datasets which attract a high number of visitors on the EU ODP and have a high number of downloads;
- Datasets whose owners proactively express to the EU ODP team their intention to collaborate.

In each case, PwC collaborated with both the data owners and the EU ODP team to propose, define and implement the visualisations. PwC provided technical support as needed and hosted the data visualisations on its pilot server for its future handover to the EU ODP servers.

All the visualisations presented below can be accessed from home pages (http://52.50.205.146:8890/data-visualisation-pilot/CORDIS/, http://52.50.205.146:8890/data-visualisation-pilot/TransparencyRegister/, http://52.50.205.146:8890/data-visualisation-pilot/TED/map/ted_map.html) hosted by PwC. These pages give the basic information about the visualisations and dataset. The page also contains links, under the form of images, to reach the pages on which the visualisations are hosted.

The licence associated with all the development done on this project (i.e. the data preparation scripts and visualisations) is the European Union Public Licence (EUPL v.1.1)[4].

This remainder of this document is structured as follows:

- Section Annex XI documents the visualisations provided on the CORDIS dataset. It contains a presentation of the dataset, details on the work realised (e.g. script developed), technologies and data used and some screenshots of the visualisations.
- Section 3 documents the visualisations provided on the Transparency Register dataset. It contains a presentation of the dataset, details on the work realised (e.g. script developed), technologies and data used and some screenshots of the visualisations.
- Section 4 documents the visualisation provided on the TED dataset. It contains a presentation of the dataset, details on the work realised (e.g. script developed), technologies and data used and a screenshot of the visualisation.

[1] https://data.europa.eu/euodp/en/data/dataset/cordisref-data

[2] https://data.europa.eu/euodp/en/data/dataset/transparency-register

[3] https://data.europa.eu/euodp/en/data/dataset/ted-1

[4] https://joinup.ec.europa.eu/community/eupl/og_page/european-union-public-licence-eupl-v11

2 CORDIS VISUALISATIONS

CORDIS is the European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their results in the broadest sense. Two types of datasets are available on CORDIS:

- **The project datasets** which list information for each EU-funded projects (e.g. start date, end date, participants' name). The main fields present in this kind of datasets can be found in Annex II - Table 2; and
- **The organisation datasets** which list information about organisations (e.g. name, role, address) for all EU-funded projects. For more information on the main fields in these datasets, please refer to Annex II - Table 3.

All the visualisations presented in this section are based on datasets from CORDIS:

- FP6: running from 2002 to 2006[5];
- FP7: running from 2007 to 2013[6]; and
- H2020: running from 2014 to 2020[7].

All the visualisations are accessible from the homepage (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/CORDIS/) by clicking on an image representing the chart. The navigation through the different pages can also be done using the tabs on top of every page (differentiating the active page form the other using different colours). These two possibilities are represented on the screenshot on Figure 1.

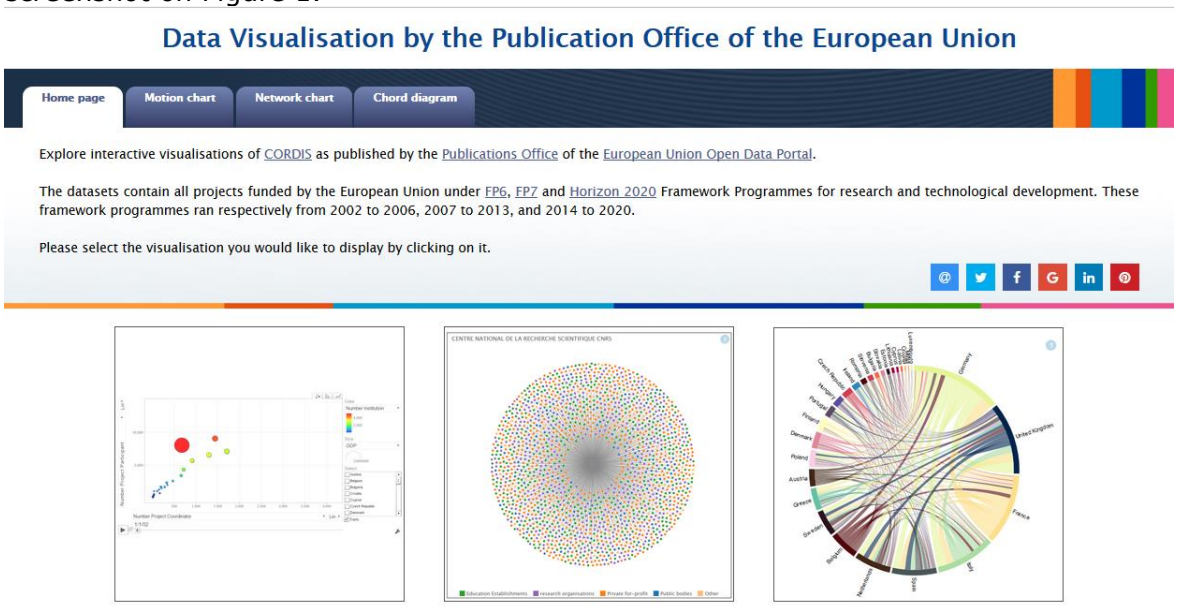


Figure 1: CORDIS home page

2.1 Motion Chart using Gapminder

The visualisation (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/CORDIS/motionChart/layout_motionChart.html) allows exploring the

[5] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[6] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[7] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects

existing relations between different dimensions for each of the EU-28 countries and the participation of these countries in EU-funded projects. The available dimensions are:

- Number of coordinated projects
- Budget of coordinated projects
- Total number of projects
- Number of projects participation as participant
- GDP
- Population
- % pop. with a secondary education
- % pop. with a tertiary education
- R&D Budget
- GDP per capita
- Budget of coordinated projects / capita
- Number of coordinated projects – Cumulative
- Budget of coordinated projects – Cumulative
- Total number of projects – Cumulative
- Number of projects as participant – Cumulative
- Budget of coordinated projects / capita – Cumulative

It is inspired by Trendalyzer, an information visualisation software for animation of statistics that was initially developed by Hans Rosling's Gapminder Foundation, now owned by Google Inc.

The data visualised have been extracted from CORDIS Datasets (framework contract FP6[8], FP7[9] and H2020[10]) and Eurostat (Population of EU Member State[11], GDP of EU Member State[12], population of educational attainment level[13] and Total intramural R&D expenditure[14]). The data used[15], the code developed for data preparation[16], the code used for the visualisation[17] as well as the data visualised[18] are available on GitHub.

The visualisation allows to analyse the data through an interactive bubble chart (cf. Figure 2).

The bubble chart shows five variables:

- Two numeric variables on the X and Y axes,
- Bubble size and colour, and
- A time variable that may be manipulated with a slider.

The software uses brushing and linking techniques to display the numeric value of a highlighted country. By default, the size and colour of the bubble are respectively determined by the GDP (Gross domestic product) of the country and the number of institutions in the country. The position of the country in the chart is determined by the number of projects with the role of coordinator versus the role of participant within the country.

[8] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[9] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[10] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[11] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en
[12] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=naida_10_gdp&lang=en
[13] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_03&lang=en
[14] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=rd_e_gerdtot&lang=en
[15] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets
[16] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep
[17] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/motionChart
[18] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets

The timeline below allows to display the evolution between the 3 Framework Programmes (FP6, FP7, H2020). All the dimensions can be updated using the respective dropdowns.

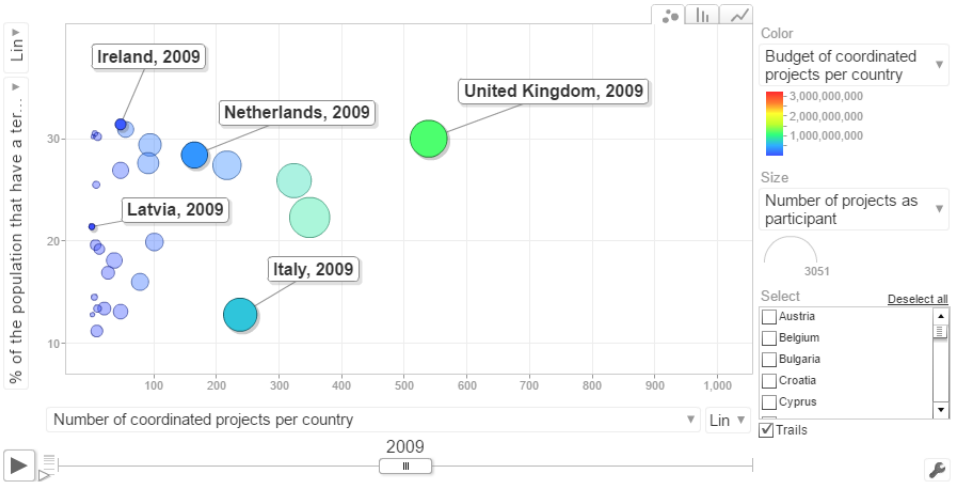


Figure 2: First screen of the motion chart - Bubble chart

The data can also be explored using a bar chart (cf. Figure 3) and a line chart (cf. Figure 4).

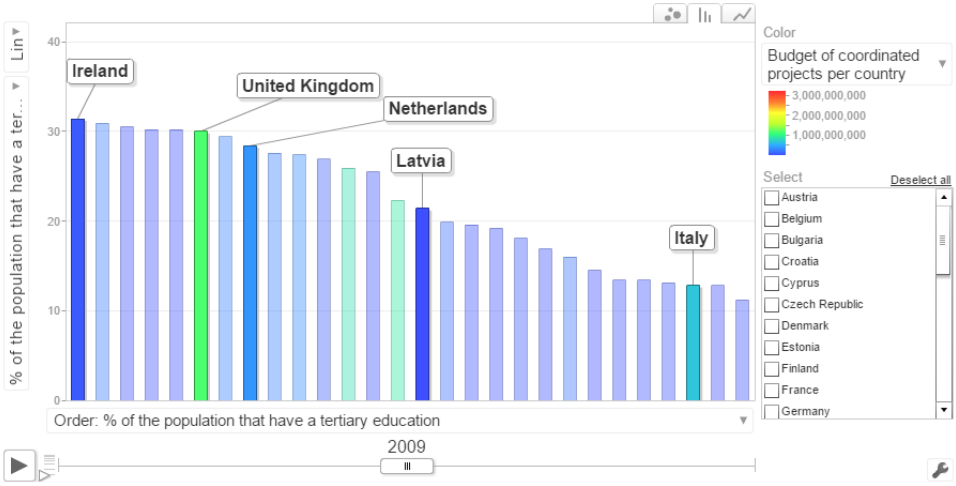


Figure 3: Second screen of the motion chart - Bar chart

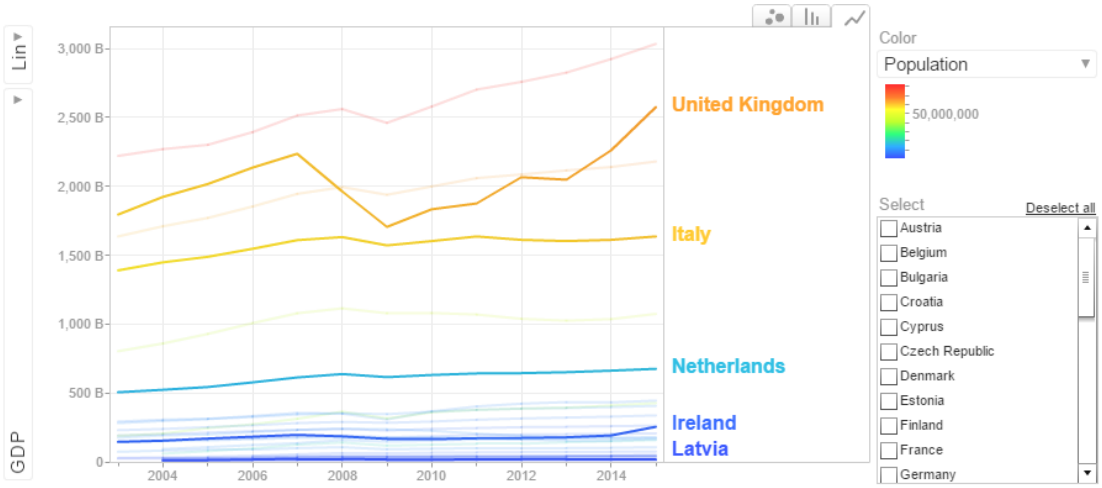


Figure 4: Third screen of the motion chart - Line chart

2.1.1 Activities

In order to provide the visualisations the following tasks have been carried out:

- Development of an R script[19] to gather and transform the data. The goal of this script is to aggregate the CORDIS datasets of the different framework programmes into one dataset and to merge this dataset with datasets from Eurostat.
- The output of the script is a JS file[20] containing, for all countries, the dimensions listed in section 2.1. The JS file generated is used to directly feed the motion chart.
- Creation of a HTML code[21] that defines the structure of the web page (see Annex IV).
- Creation of a JavaScript script (include in the HTML) using the Google-visualization library[22] to:
    - Query the JS and retrieve the data;
    - Configure the chart; and
    - Draw the chart.
- Creation of CSS rules to design the layout and style of the page.

2.1.2 Server

Two HTML files containing the HTML, JavaScript and CSS code (one for the iframe and one containing the entire page) were uploaded on the server, in the sub-folder "motionChart" contained in the "CORDIS" folder. Furthermore, the motionChart.js file containing the data used to feed the chart has been uploaded to the "Dataset" folder.

Finally a link was added from the CORDIS homepage that allows to access the chart by clicking on an image representing the motion chart.

2.1.3 Configuration

In order to enhance the readability and, on request by the client, the speed to which the animation of the chart is "played" has been set to the minimum.

[19] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/DataPrep/dataPrep_motionChart.R

[20] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/Datasets/motionChart.js

[21] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/motionChart

[22] https://developers.google.com/chart/interactive/docs/

2.2 Network Diagram by organisation

The visualisation (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/CORDIS/orgNetwork/layout_network_organisations.html) allows exploring the partners of EU funded research projects and their relationships.

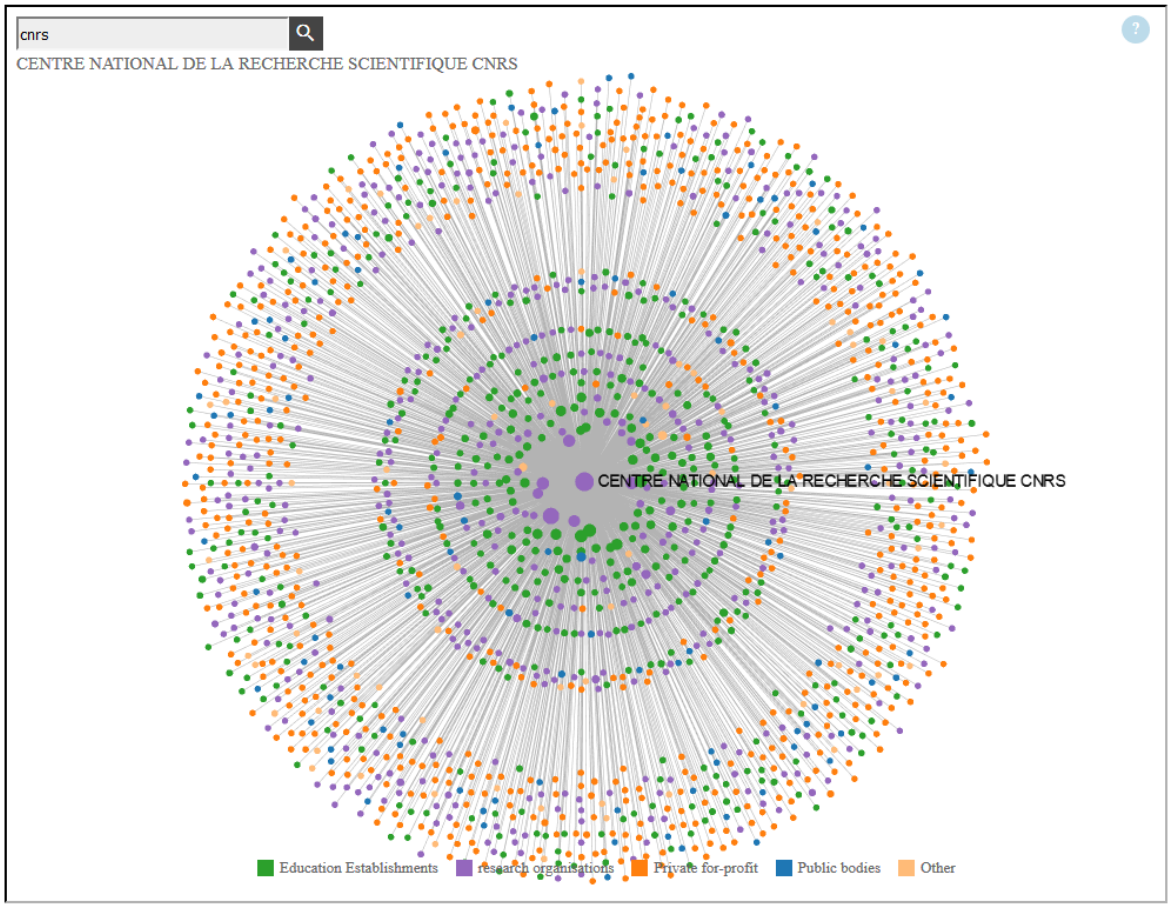


Figure 5: Network diagram per organisation

For one organisation:

- The size of the bubble represents the number of projects in which the organisation is participating;
- Its colour represents the activity of the organisation (e.g. Research Organisations or Private for-profit) and
- The length of the link represents the strength of the collaboration (i.e. the number of projects run in common by two organisations).

The selection of the organisation can be done through the filter on the top left of the organisation. Filters can be applied on activities types and countries.

The data visualised are been extracted from CORDIS Datasets (framework contract H2020[23]). The data used[24], the code developed for data preparation[25] and the visualisation[26] are available on GitHub.

Only the projects funded under H2020 have been visualised. The decision was made not to visualise the projects funded under FP6 and FP7 as the field "activityType", which

[23] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[24] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets/inputData
[25] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep
[26] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/orgNetwork

describes the type of activity carried out by the organisation, is not filled in for 18% of the organisation in FP7 and 99% of the organisations in FP6.

2.2.1 Activities

In order to provide the visualisations the following tasks were carried out:

- An R script[27] was developed to gather and transform the data. The goal of this script was, for each of the organisations, to find the other organisations it works with and the amount of programs they are working together on. The output of the script is two CSV files per organisation[28] containing the needed information.
- A python script was created to automatically generate the list of country name as HTML elements.
- A Java script[29] was created and transformed to an executable JAR file[30]. This scrip reads the CSV files and transforms them, through gephi, to gexf files formatted as needed to feed the network chart. One gexf file is available per organisation.
- A HTML code[31] defining the structure of the web page (see Annex V) was created.
- A JavaScript[32] code was adapted from the D3 example[33] to:
    - Read the gefx to get the data;
    - Configure the chart;
    - Configure the dropdowns selection; and
    - Draw the chart.
- Creation of a CSS file to design the layout of the page.

2.2.2 Server

The HTML, JavaScript and CSS files containing the code were uploaded on the PwC server. The main folder contains the two html files, the gefx files are included in the js subfolder, and the css files are included in the css subfolder. Finally, a link was added from the data-visualisation-pilot homepage to access the chart from it by clicking an image representing this network chart.

2.2.3 Configuration done

No particular configuration was needed in the context of this visualisation.

2.3 Chord Diagram

The visualisation (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/CORDIS/chordDiagram/layout_chord.html) allows exploring the collaboration between EU-28 countries on projects funded under the FP6 (2002-2006), FP7 (2007-2013), and H2020 (2014-2020) Framework Programmes:

- The countries are arranged radially around a circle;
- The links between the countries represent the number of projects that include at least one organisation of these two countries;
- The size of the arc depends on the number of projects in which the country participated.; and
- The size of the link depends on the number of projects between the two linked countries.

[27] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep

[28] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets/outputOrgNetwork/

[29] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep/organisationsNetwork%20workspace/orgNetwork/

[30] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/DataPrep/dataPrep_orgNetwork.jar

[31] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/orgNetwork/

[32] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/orgNetwork/js/

[33] https://github.com/paulovn/movie-network

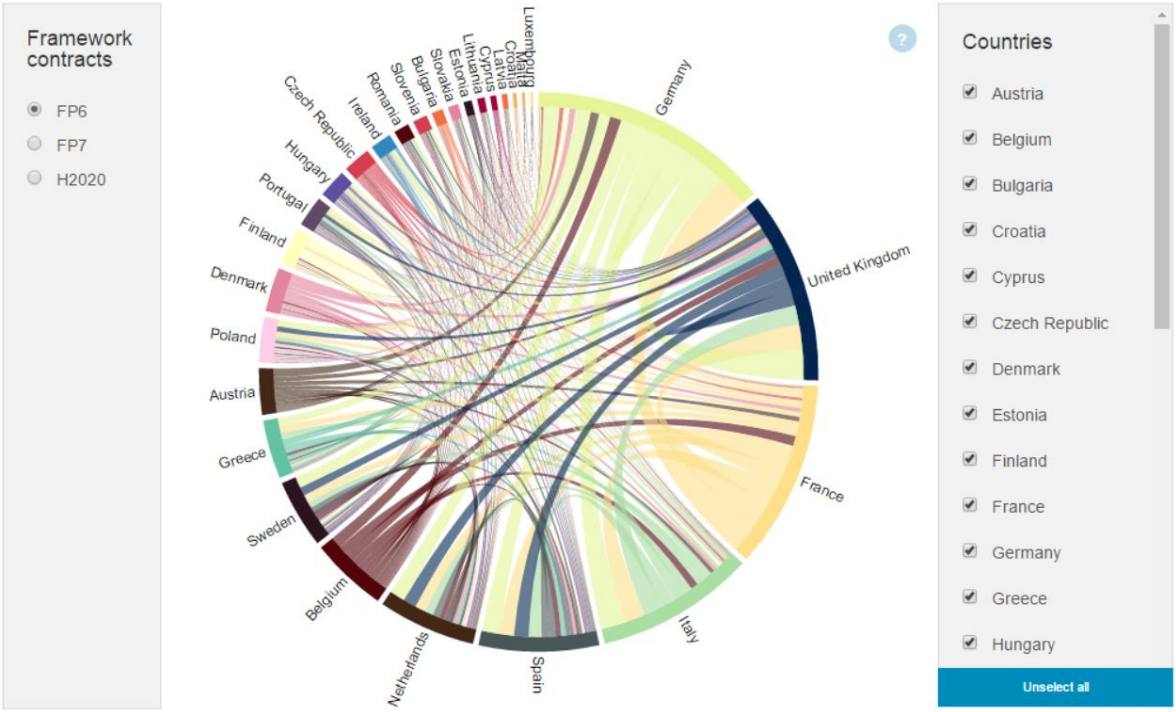


Figure 6: Chord diagram

The data visualised has been extracted from CORDIS Datasets (framework contract FP6[34], FP7[35] and H2020[36]). The data used[37], the code developed for data preparation[38], the code used for the visualisation[39] as well as the data visualised[40] are available for reuse on GitHub.

2.3.1 Activities

In order to provide the visualisation the following tasks were carried out:

- Gathering and transformation of the data: a CSV file was created through an R[41] script. For each of the organisation dataset (FP6, FP7 and H2020) the following logic is applied:
    - Select the organisations that are in one of the EU-28 country;
    - Group the organisations per project;
    - Keep only one organisation per country;
    - Create a matrix with the number of projects that link two countries.
- Creation of a HTML code[42] that defines the structure of the web page (see Annex VI).
- Creation of a JavaScript script[43] using the D3.js library to:
    - Read the data;
    - Configure the chart;

[34] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[35] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[36] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[37] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets
[38] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep
[39] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chordDiagram
[40] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets
[41] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/DataPrep
[42] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chordDiagram
[43] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chordDiagram/js

    - Configure the filters behaviour and add a button to select/unselect all countries;
    - Configure the chart behaviour: highlighting and greying out of the country links.
    - Draw the chart;
    - Help bubble displaying a text explaining how to use the filters and chart interactivity features.
- Creation of a CSS to design the layout of the page.

2.3.2 Server

The HTML, JavaScript and CSS files containing the code were uploaded on the PwC server. The main folder is the "chordDiagram" inside the "CORDIS" folder, containing sub-folders. The "css" subfolder contains the CSS files defining the styles. The "js" subfolder contains the JavaScript scripts and the "chart" subfolder contains the html file. The other subfolders contain libraries which should not be modified. The file chord.csv containing the data is in the "Dataset" folder inside the "CORDIS" folder.

Finally, a link was added from the CORDIS homepage to access the chart from it by clicking an image representing this chord diagram.

2.3.3 Configuration done

The thicker the lines between two countries, the more projects they have worked on together under the selected framework programme, and vice versa.

Two filters were created: one to select the framework contract and another to highlight/grey out countries.

2.4 Update of the visualisations

To update the visualisations, version 3.3.2 of R needs to be installed, please refer to Annex III, I.1 for an installation guide of R. If the network diagram needs to be updated, Java SE Runtime Environment 8 is needed, please refer to Annex III, I.3 for an installation guide.

Second, the most up to date content of the GitHub project needs to be downloaded. Please refer to Annex III, I.2 for step-by step instructions.

To update the data, open the command prompt on your computer from the CORDIS\DataPrep folder in the GitHub project and execute the following command with one of the following as argument:

1. -A or --All to update all visualisations, followed by the year until which the data should be updated;
2. -M or –-MotionChart to update only the motion chart, followed by the year until which the data should be updated;
3. -O or –-OrgNetwork to update only the motion chart;
4. -C or –-Chord to update only the chord diagram;
5. -U or –-Update to download the latest CORDIS data;
6. -h or –-help to show display the help;
7. Or any combination of point 2, 3 or 4.

```
"C:\Program Files\R\R-3.3.2\bin\Rscript.exe" "dataPrep_CORDIS.R" argument
```

Wait for the script to finish, this can take up to a few minutes.

The updated files will be created in the CORDIS\Datasets folder of the GitHub project as following:

- Chord.csv[44] will be updated when updating the chord diagram. Please upload it to the CORDIS\Datasets folder on the server.
- motionChart.js[45] will be updated when updating the motion chart. Please upload it to the CORDIS\Datasets folder on the server.
- organisations.js[46] and the content of the outputOrgNetwork subfolder will be updated when updating the organisations network. Please upload he organisations.js file to the CORDIS\Datasets folder on the server. The files in the outputOrgNetwork subfolder are an interim result, they need further processing as explained in the following paragraph.

To update the network diagram, a second script needs to be executed. Open the command prompt on your computer from the CORDIS\DataPrep folder in the GitHub project. Type java -version into your command line. If Java is installed, you will see a message stating what version of Java is currently installed. Make sure this is version "1.8.xxx".

Next type the command below with the following two arguments:

1. The path to the input folder with the files that need to be processed. By default it is CORDIS\Datasets\outputOrgNetwork.
2. The path to the output folder to which the processed files should be written.

```
java –jar dataPrep_orgNetwork.jar inputFolder outputFolder
```

Important to note is that this program takes several hours to finish. The program will alert you each time it has processed 100 files and will give a list of all files that were not processed correctly at the end.

After the program is finished, upload all files created in the folder you specified as the output folder to the CORDIS\Datasets\orgNetwork folder on the server.

[44] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/Datasets/Chord.csv
[45] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/Datasets/motionChart.js
[46] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/Datasets/organisations.js

3 TRANSPARENCY REGISTER DATASET

The Transparency Register was created and is operated by both the European Parliament and the European Commission. The goal of this initiative is to make the EU decision-making process more transparent and open, in order to ensure a balanced representation and to avoid illegitimate or privileged access to information. Moreover, the idea of a more transparent and open process is to allow and encourage citizens to take part in the democratic life of the EU. This dataset groups information about the interests pursued for every institutions and the budgets related to these activities.

All the visualisations created are accessible from the homepage (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/TransparencyRegister/) by clicking on an image representing the chart. The navigation through the different pages can also be done using the tabs on top of every page (differentiating the active page form the other using different colours). These two possibilities are represented on the screenshot on Figure 7.

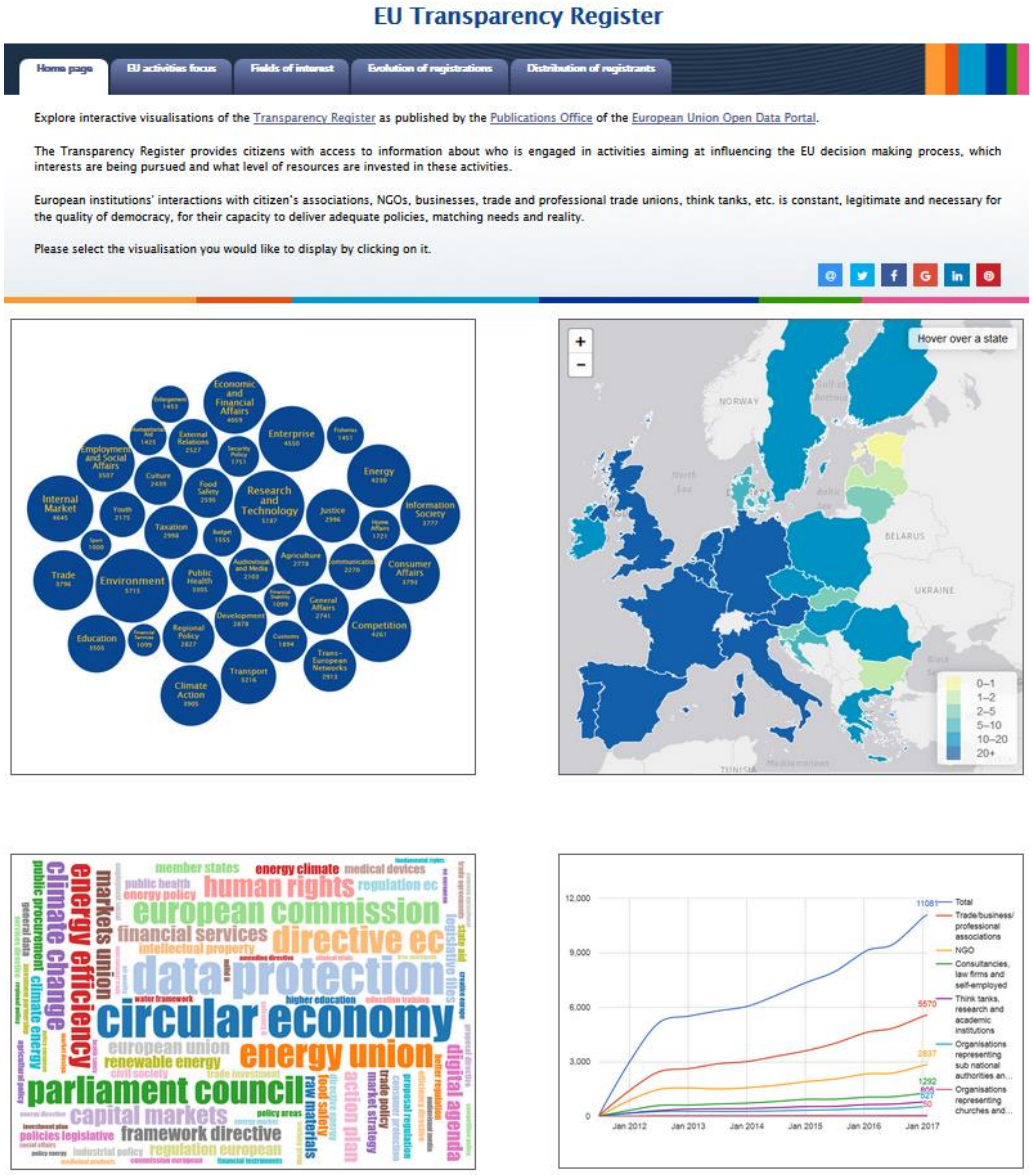


Figure 7: Transparency Register homepage

3.1 Bubble chart

The visualisation (accessible at the following URL http://data-visualisation.semic.eu:8890/TransparencyRegister/bubbleChart/layout_bubble.html) allows to see which interests are pursued by the organisations.

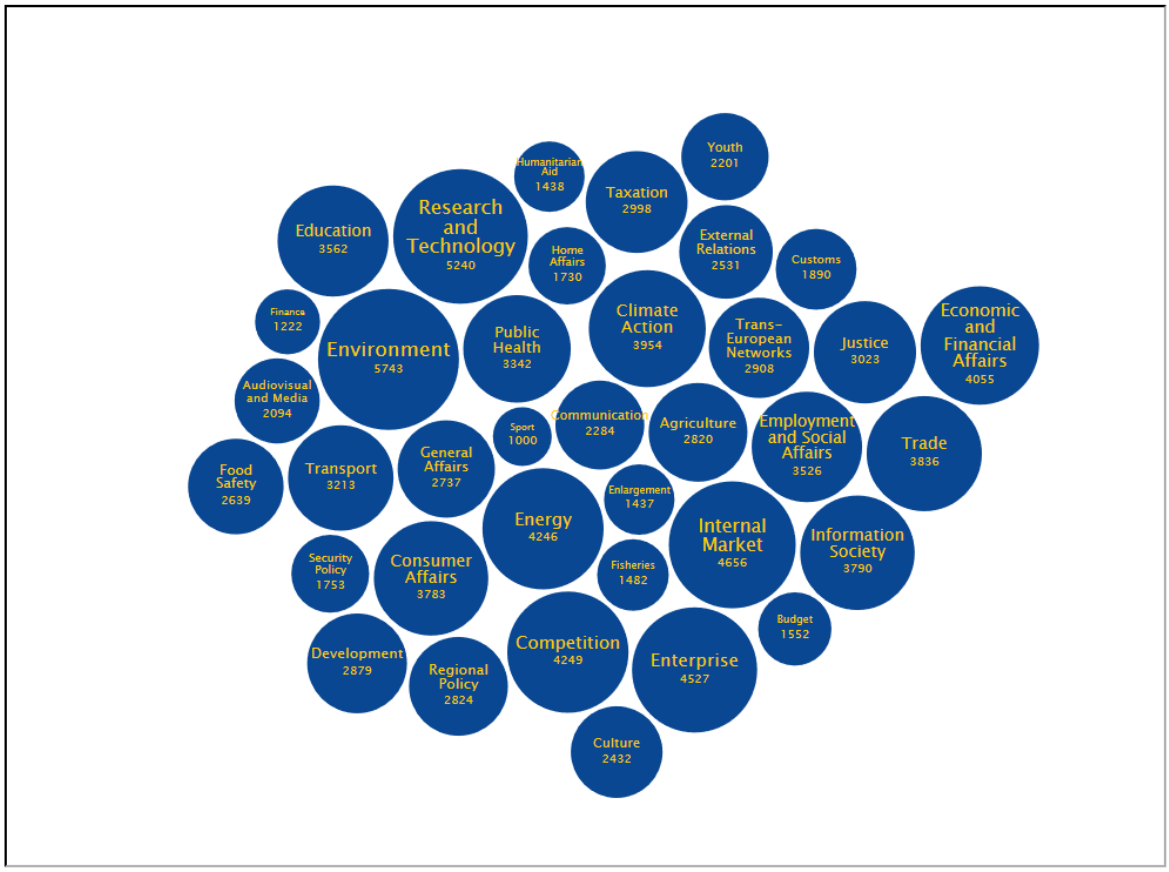


Figure 8: Bubble chart

Every bubble represents a domain and its size directly depends on the number of entities interested in this specific domain. The number of entities is also indicated for each bubble by the number below the domain name.

3.1.1 Activities

In order to provide the visualisations the following tasks have been carried out:

- An R script[47] was developed to gather and transform the data. The output of the script was a CSV file[48] containing the aggregated data per domain.
- Creation of a HTML code[49] that defines the structure of the web page (see Annex VII).
- Creation of a JavaScript script[50] using the D3.js library to:

[47] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/dataPrep/DataPreparation.R

[48] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/BubbleChartData.txt

[49] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/bubbleChart

[50] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/bubbleChart/bubble_force.js

- Read the data from the CSV;
- Configure and draw each bubble and their container.

- Creation of CSS[51] rules to design the layout and style of the page.

3.1.2 Server

The files containing the HTML, JavaScript and CSS code were uploaded on a PwC server and placed in the "bubbleChart" subfolder of the "TransparencyRegister" folder and the text file containing the data in CSV format has been uploaded to the "Datasets" subfolder. Moreover a link was added from the TransparencyRegister homepage that allows to access the chart by clicking on an image representing the bubble chart.

3.1.3 Configuration

The bubble chart is interactive in the sense that the bubble can be moved and dragged.

3.2 Word cloud

The visualisation (accessible at the following URL http://data-visualisation.semic.eu:8890/TransparencyRegister/wordCloud/layout_wordcloud.html) allows to visualise the importance of EU field activities.

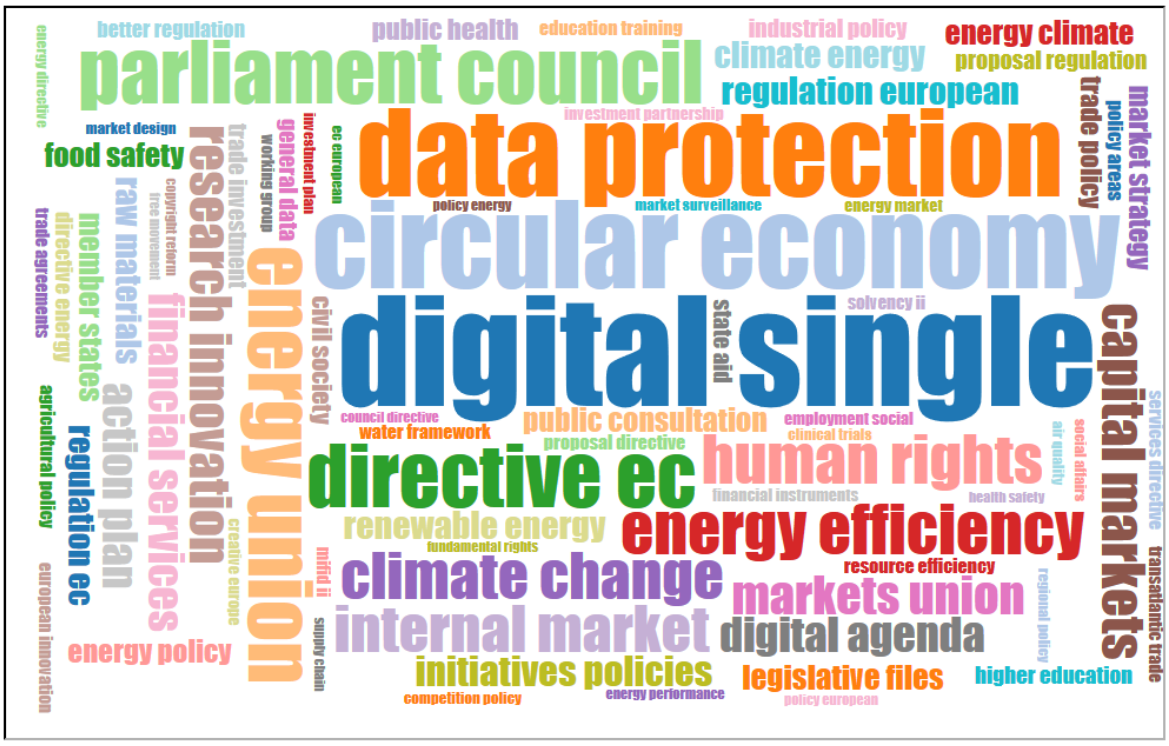


Figure 9: Word cloud

Every group of words represents a field activity and its size depends on the number of times it appears in the dataset.

3.2.1 Activities

In order to provide the visualisations the following tasks have been carried out:

[51] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/bubbleChart/style.css

- An R script[52] was developed to gather and transform the data. First, the texts are cleaned: the stop words are removed (for several languages) as well as the special characters like the spaces and punctuation. The output of the script is a JS file[53] containing the aggregated data under the form of a matrix containing the words and their number of occurrences.
- Creation of a HTML code[54] that defines the structure of the web page (see Annex VIII).
- Creation of the JavaScript script (included in the HTML) using the D3.js library to configure and draw the visualisation.
- Creation of CSS rules (included in the HTML) to design the layout and style of the page.

3.2.2 Server

The files containing the HTML, JavaScript and CSS code were uploaded on a PwC server and placed in the "wordCloud" subfolder of the "TransparencyRegister" folder and the JS file containing the data in the "Datasets" subfolder.

Moreover a link was added from the "TransparencyRegister" homepage that allows to access the chart by clicking on an image representing the word cloud.

3.2.3 Configuration

It has been decided to use to only use groups of two words and to remove words smaller than two characters.

3.3 Line chart

The visualisation (accessible at the following URL http://data-visualisation.semic.eu:8890/TransparencyRegister/lineChart/layout_linechart.html) allows to visualise the evolution of the number of organisation for every type of organisation.

Every line represents the evolution for a type of organisation. The blue line on top represents the evolution for the total number of organisations.

[52] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/dataPrep/DataPreparation.R
[53] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/wordCountMapping_activity_twogram_multiplied_100.js
[54] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/wordCloud

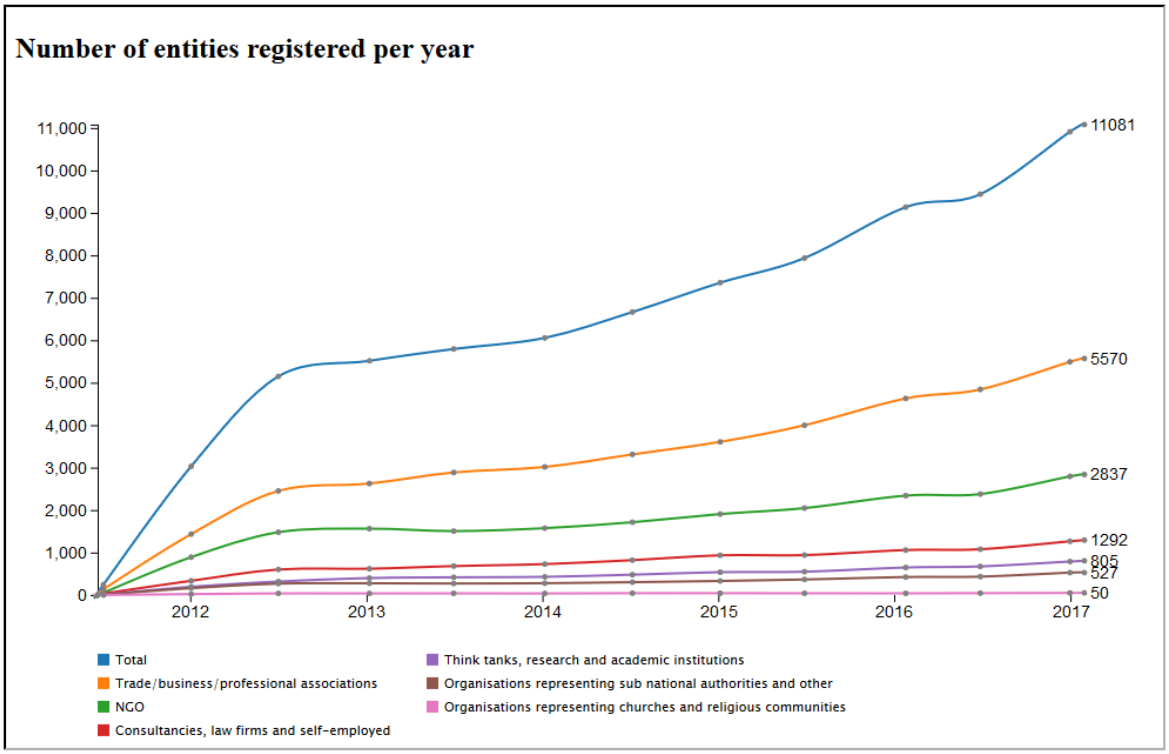


Figure 10: Line chart

3.3.1 Activities

In order to provide the visualisations the following tasks have been carried out:

- Creation of a R script[55] to aggregate the data for every type of organisation in a CSV file[56].
- Creation of the JavaScript script[57] (included in the HTML) using the D3.js library to:
  - Read the data in the CSV file;
  - Configure and draw the visualisation.
- Creation of a HTML code[58] that defines the structure of the web page (see Annex VIII).

3.3.2 Server

The files containing the HTML, JavaScript and CSS code were uploaded on a PwC server and placed in the "lineChart" subfolder of the "TransparencyRegister" folder and the CSV file containing the data was put in the "Datasets" subfolder.

Moreover a link was added from the TransparencyRegister homepage that allows to access the chart by clicking on an image representing the line chart.

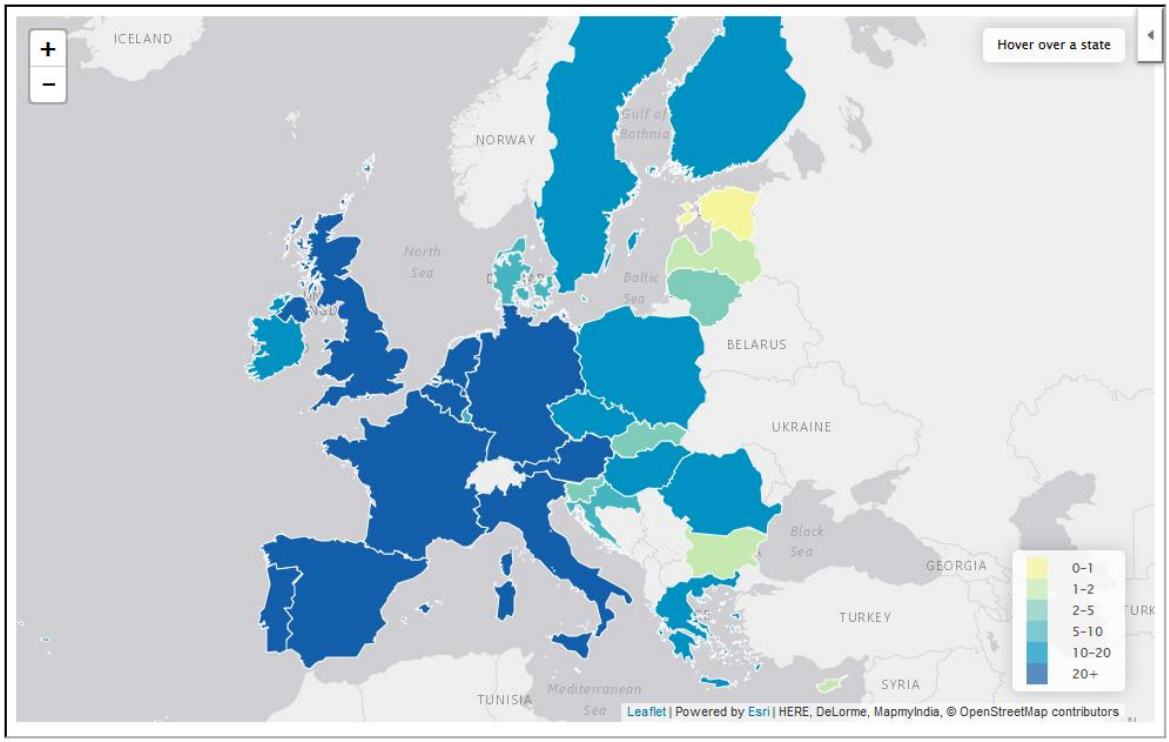
3.3.3 Configuration

An additional line has been drawn by aggregating the data in order to represent the evolution of the total number of organisation. Moreover, labels have been placed next to each line to precisely state the value for the most recent value.

[55] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/dataPrep/DataPreparation.R
[56] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/Data_linechart.csv
[57] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/lineChart
[58] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/lineChart

3.4 Map

The visualisation (accessible at the following URL http://data-visualisation.semic.eu:8890/TransparencyRegister/map/layout_leafletMap.htmlhttp://data-visualisation.semic.eu:8890/TransparencyRegister/layout_wordcloud.html) presents the number of entities by country for every category or subcategory.



Every country is coloured according on the number of entities belonging to the selected (sub)category from the country. The filter on the right allow to select a category or subcategory. Label on the top-right corner display the exact number for a specific country when it is hoovered.

3.4.1 Activities

In order to provide the visualisations the following tasks have been carried out:

- Creation of an R script[59] to aggregate the data for every category and subcategory.
- Creation of a HTML code[60] that defines the structure of the web page (see Annex X).
- Creation of a JavaScript files[61] to store the data in an adequate format to be used by the visualisation (values and geographical information to draw the borders).
- Creation of the JavaScript script (included in the HTML) using the leaflet library to:
    - Configure and draw the visualisation.

3.4.2 Server

The files containing the HTML, JavaScript and CSS code were uploaded on a PwC server and placed at the root of the TransparencyRegister folder as well as the CSV file containing the data.

[59] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/dataPrep/DataPreparation.R

[60] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/map

[61] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/TransparencyRegister/Datasets

Moreover a link was added from the data-visualisation-pilot/TransparencyRegister homepage that allows to access the chart by clicking on an image representing the map chart.

3.5 Update of the visualisations

To update the visualisations, version 3.3.2 of R needs to be installed, please refer to Annex III, I.1 for an installation guide of R.

Secondly, the most up to date content of the GitHub project needs to be downloaded. Please refer to Annex III, I.2 for step-by step instructions.

Open the command prompt on your computer from the TransparencyRegister\dataPrep folder in the GitHub project.

Finally, execute below command, no arguments are required:

```
"C:\Program Files\R\R-3.3.2\bin\Rscript.exe" "DataPreparation.R"
```

Wait a moment for the script to finish, this can take up to a few minutes.

The updated files will be created in the TransparencyRegister\Datasets folder of the GitHub project as following:

- Full_export_new.xml[62] contains raw data and will be updated, all subsequent files will be updated based on this file and the Countries.csv file. The Countries.csv file does not need to be updated;
- datasetMap.js[63] will be updated, this file contains the information for the map. The map also needs the eu-states.js file which contains the borders of all countries, this file does not need to be updated. Please upload the datasetMap.js file to the TransparencyRegister\Datasets folder on the server;
- BubbleChartData.txt[64] will be updated, this file contains the information for the bubble chart. Please upload this file to the TransparencyRegister\Datasets folder on the server;
- wordCountMapping_activity_twogram_multiplied_100.js[65] will be updated, this file contains the information for the word cloud. Please upload this file to the TransparencyRegister\Datasets folder on the server;
- Data_linechart.csv[66] will be updated, this file contains the information for the line chart. Please upload this file to the TransparencyRegister\Datasets folder on the server.

[62] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/full_export_new.xml

[63] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/datasetMap.js

[64] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/BubbleChartData.txt

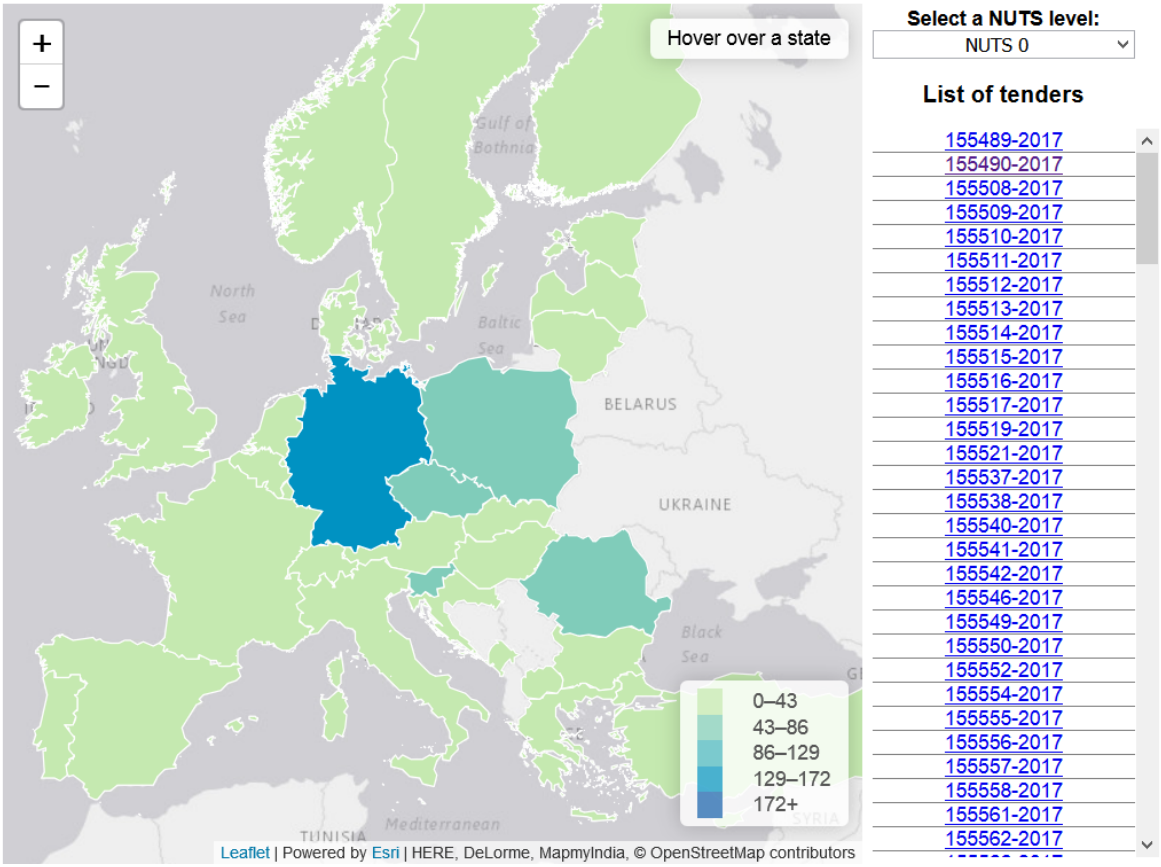
[65] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/wordCountMapping_activity_twogram_multiplied_100.js

[66] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TransparencyRegister/Datasets/Data_linechart.csv

4 TENDERS ELECTRONIC DAILY

4.1 Map

The visualisation allows to visualise the number of tenders by NUTS (level 0, 1, 2 or 3) and to access the list of tenders by clicking a NUTS. Another click on one of the tender allows to visit the webpage related to it. The demo version is accessible at the following URL: http://data-visualisation.semic.eu:8890/TED/map/ted_map.html.



Every NUTS is coloured according on the number of tenders. The filter on the right allow to select a NUTS level. Label on the top-right corner display the exact number for a specific NUTS when it is hoovered and the list on the right display the tenders when a NUTS is clicked.

4.1.1 Activities

In order to provide the visualisations the following tasks have been carried out:

- Creation of a R script[67] to aggregate the data for every category and subcategory using the TED API[68] to retrieve the fields "ND" and "NUTS".
- Creation of a HTML code[69] that defines the structure of the different elements composing the chart.
- Creation of a JavaScript file[70] to store the data in an adequate format to be used by the visualisation (values and geographical information to draw the borders).

[67] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TED/dataPrep/dataPrep.R
[68] http://ted.europa.eu/api/swagger-ui.html
[69] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TED/map/ted_map.html
[70] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TED/datasets/TEDdata.js

- Creation of the JavaScript script (included in the HTML) using the leaflet library to:
    - Configure and draw the visualisation.
- Creation of CSS rules (included in the HTML) to design the layout and style of the page.

4.1.2 Server

The files containing the HTML, JavaScript and CSS code were uploaded on a PwC server and placed in the "map' subfolder of "TED" folder and the file containing the data was put in the "Datasets" subfolder.

4.2 Update of the visualisation

To update the visualisations, version 3.3.2 of R needs to be installed, please refer to Annex III, I.1 for an installation guide of R.

Secondly, the most up to date content of the GitHub project needs to be downloaded. Please refer to Annex III, I.2 for step-by step instructions.

Open the command prompt on your computer from the TED\dataPrep folder in the GitHub project.

Finally, execute below command, one argument is required, more specifically your key for the TED API:

```
"C:\Program Files\R\R-3.3.2\bin\Rscript.exe" "dataPrep.R" APIkey
```

Wait a moment for the script to finish, this can take up to a few minutes.

The updated files will be created in the TED\datasets folder of the GitHub project as following:

- APIoutput.js[71] will be updated, this file will be used to update the TEDdata.js file;
- TEDdata.js[72] will be updated. Please upload this file to the TED\Datasets folder on the server.

[71] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TED/datasets/APIoutput.js
[72] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/TED/datasets/TEDdata.js

Annex I. CORDIS – Organisation activity type

Table 1: list of organisation activity type

Code	Title
HES	Higher or Secondary Education Establishments
REC	Research Organisations
PRC	Private for-profit entities (excluding Higher or Secondary Education Establishments)
PUB	Public bodies (excluding Research Organisations and Secondary or Higher Education Establishments)
OTH	Other

Annex II. CORDIS – Main Fields

Table 2: Main fields of project dataset

Field	Description
Rcn	Unique ID of a project
acronym	Acronym of the project
title	Name of the project
startDate	Start date of the project
endDate	End date of the project
totalCost	Budget of the project
coordinator	Name of the coordinator organisation
coordinatorCountry	Country of origin of the coordinator organisation
participants	Names of the participants' organisations
participantCountries	Countries of the participants' organisations

Table 3: Main fields of organisation dataset

Field	Description
Rcn	Unique ID of a project
Role	Role of the organisation in this project (coordinator or participant)
Name	Name of the organisation
Activity Type	Activity of the organisation (for more information cf. **Error! Reference source not ound.**)
Address (including country)	Address of the organisation

Annex III. How to guides

I.1 Installation of R

The first step is to download the version 3.3.2 of R through the following link using a web browser of your choosing (e.g. Internet Explorer, Firefox, Google Chrome): https://cran.r-project.org/bin/windows/base/ . Click on "Download R.3.3.2 for Windows" to download the program, please refer to Figure 6 for a visual representation of the webpage containing the link to click.

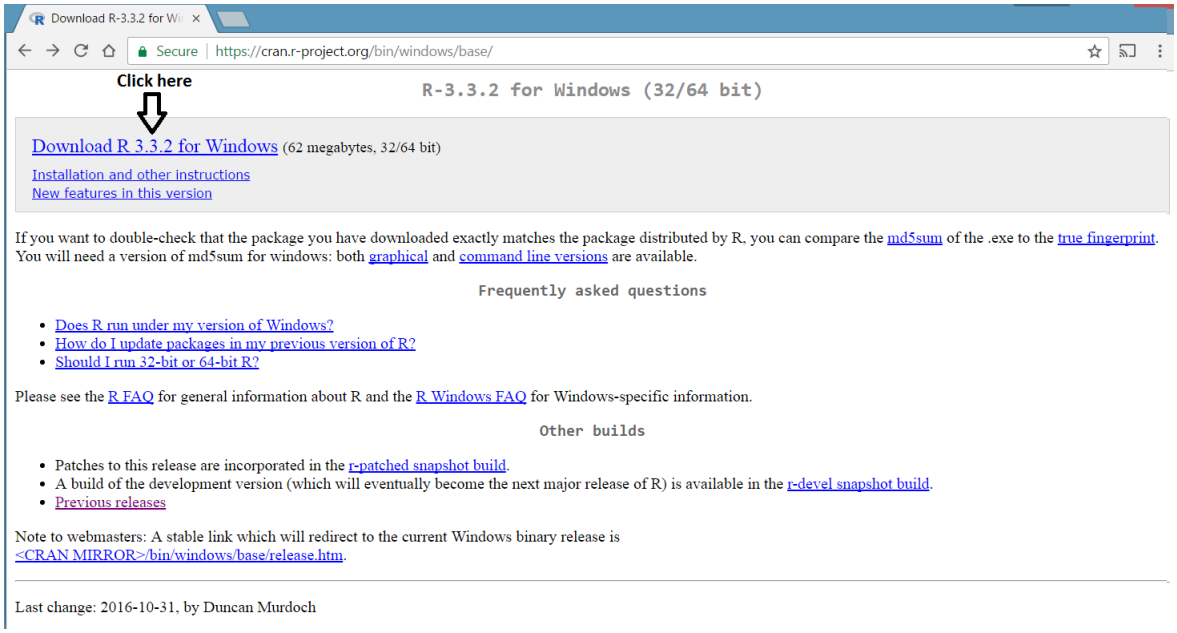


Figure 11: installation page R

If the version available here is a higher version, please find and click the version 3.3.2 on the following page: https://cran.r-project.org/bin/windows/base/old/, please refer to Figure 7 for a visual representation of this page. Subsequently, click on "Download R.3.3.2 for Windows" to download the program, please refer to Figure 8 for a visual representation of the link to click.

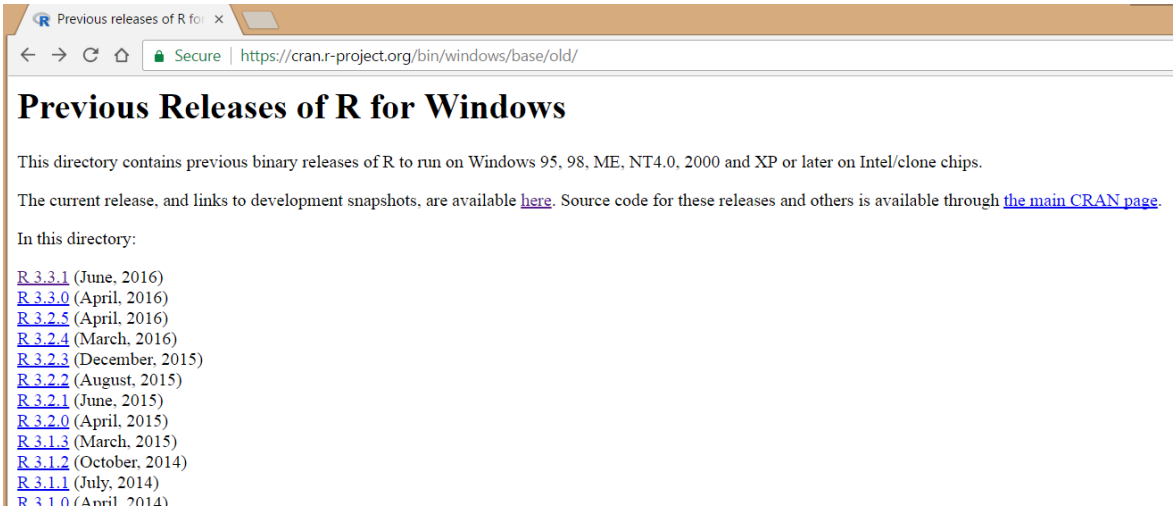


Figure 12: previous releases of R

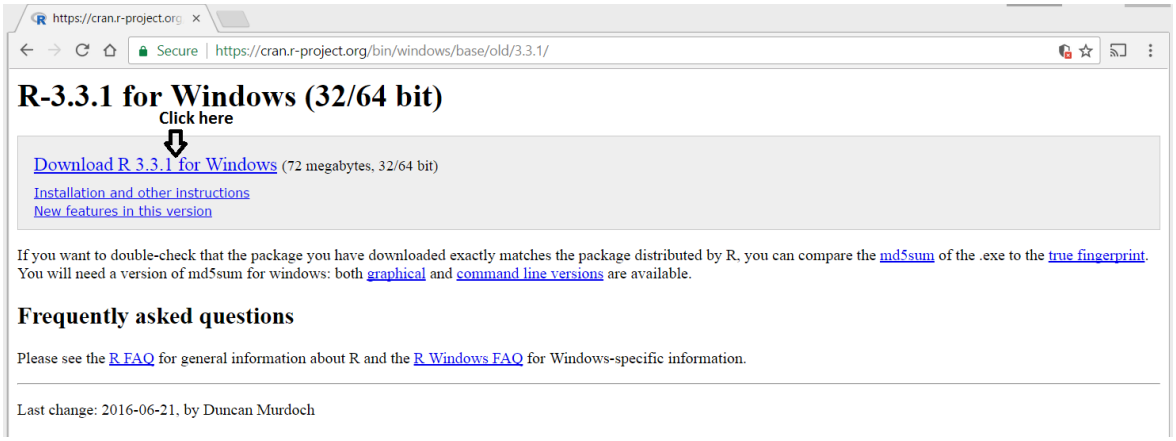


Figure 13: installation page of previous releases of R

After clicking the "Download R.3.3.2 for Windows", the installation file will download. Go to the local folder on your computer where the installer has been downloaded (by default, the "Downloads" folder) and double click the file. Refer to figure 9 for a visual representation of this.

The following screen will open (see figure 10), click OK.

Click the "next" button on all following screens (Figure 11 to 18). Please make sure that the path as displayed in Figure 13 is C:\Program Files\E\R.3.3.2.

The following screen (Figure 19) will appear, wait for the installation process to complete.

Once the installation is completed, the following screen (Figure 20) will appear. Here, click "finish".

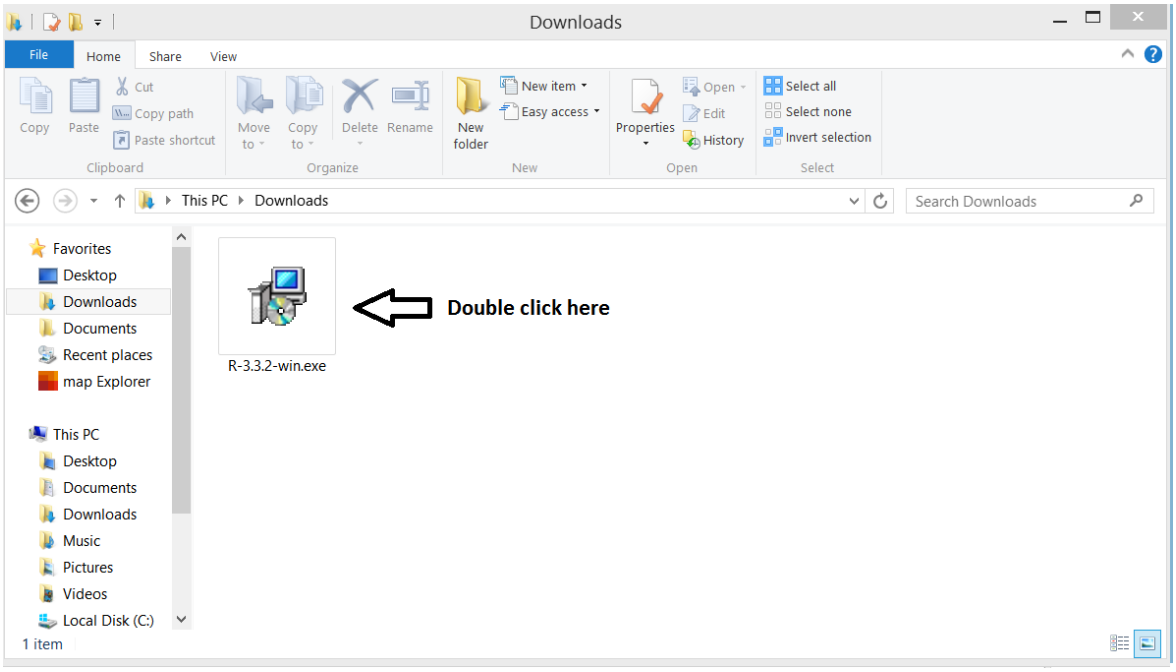


Figure 14: opening the downloaded installation file for R

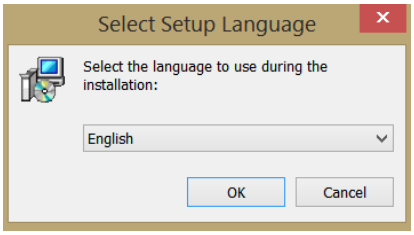


Figure 15: selection of installation language

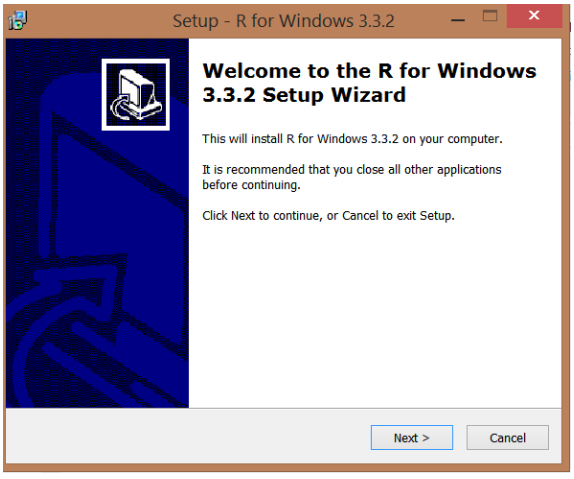


Figure 16: installation screen R (1)

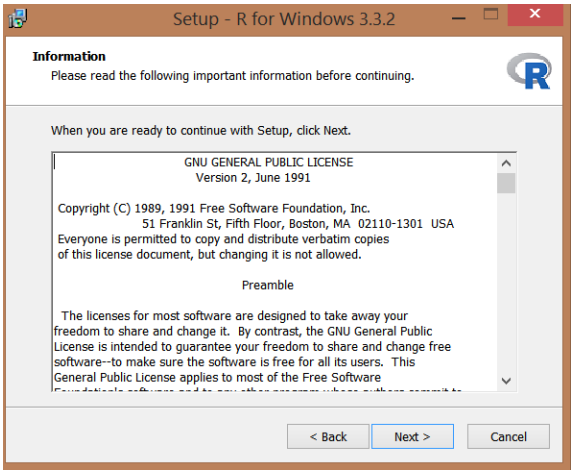


Figure 17: installation screen R (2)

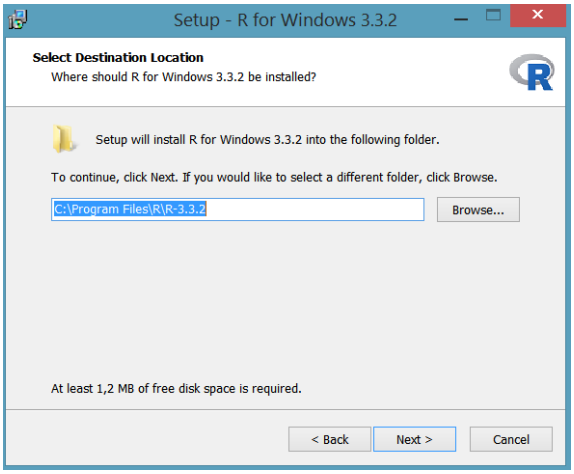


Figure 18: installation screen R (3)

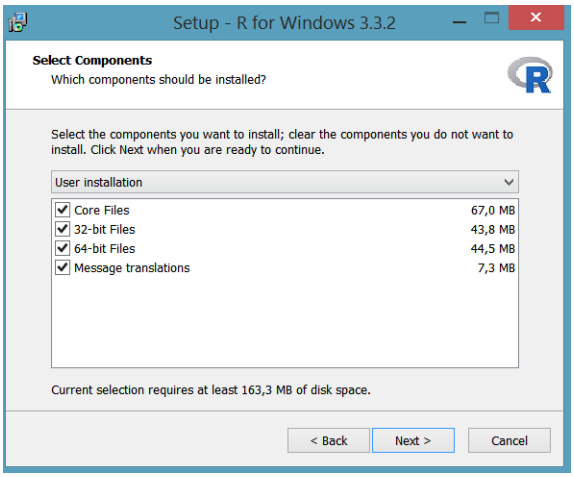


Figure 19: installation screen R (4)

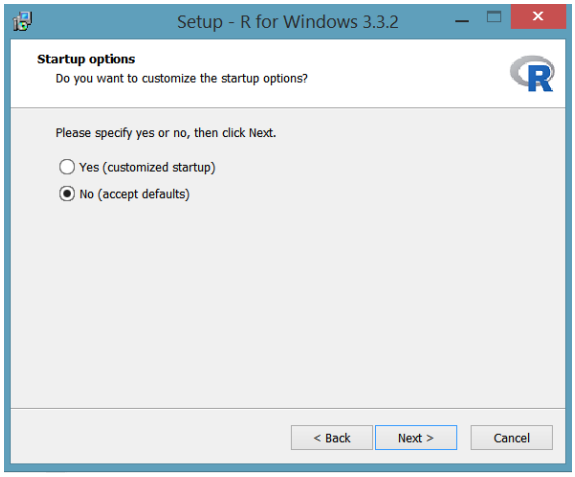


Figure 20: installation screen R (5)

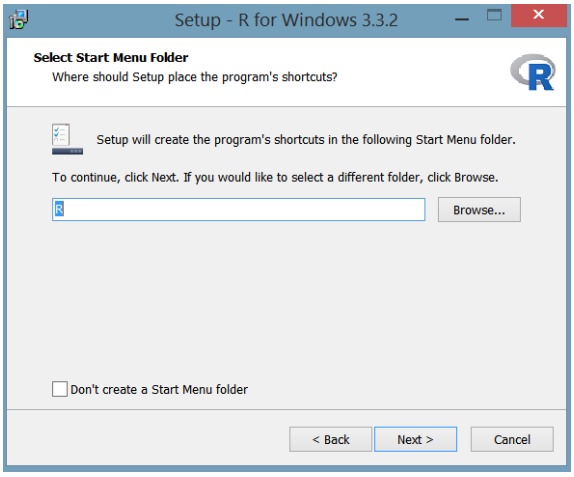


Figure 21: installation screen R (6)

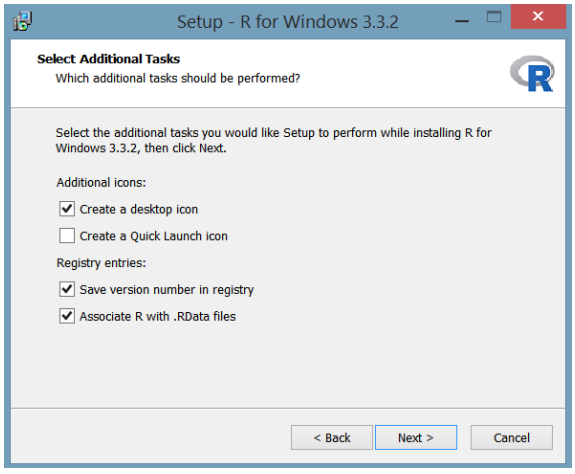


Figure 22: installation screen R (7)

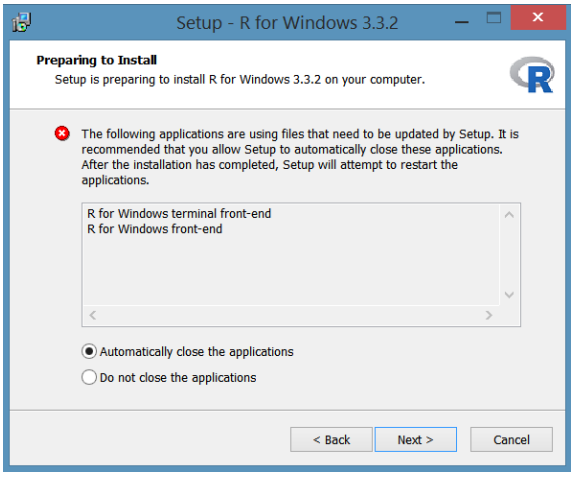


Figure 23: installation screen R (8)

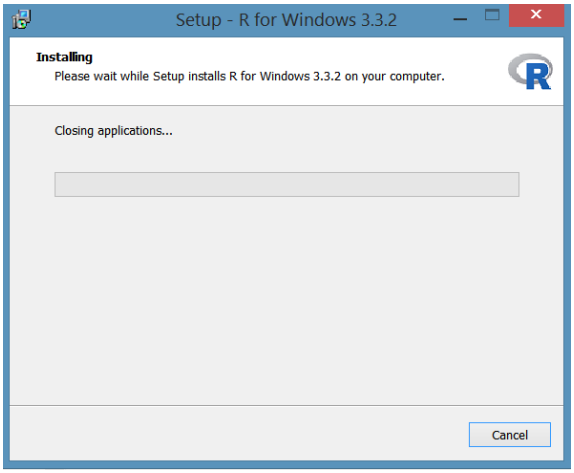


Figure 24: installation screen R (9)

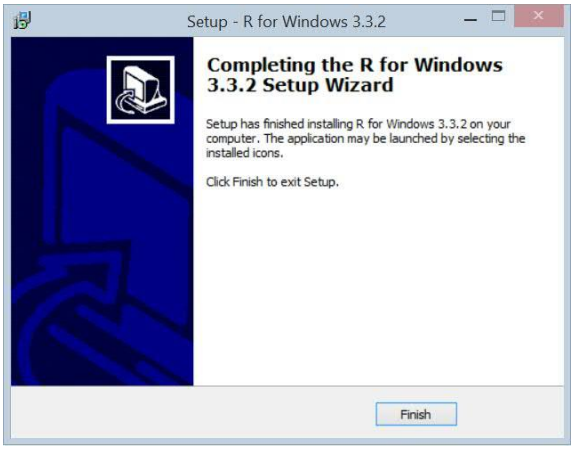


Figure 25: installation screen R (10)

I.2 Downloading the GitHub project

Download the content of the GitHub repository located at the following URL: https://github.com/SEMICeu/data-visualisation-pilot.

Click the green button "Clone or download" and select the option "Download ZIP" in the dialog that appears (see Figure 21).

This step will allow you to download the content of the GitHub project in a zip archive. Unzip it in a folder of your choosing on your hard drive.

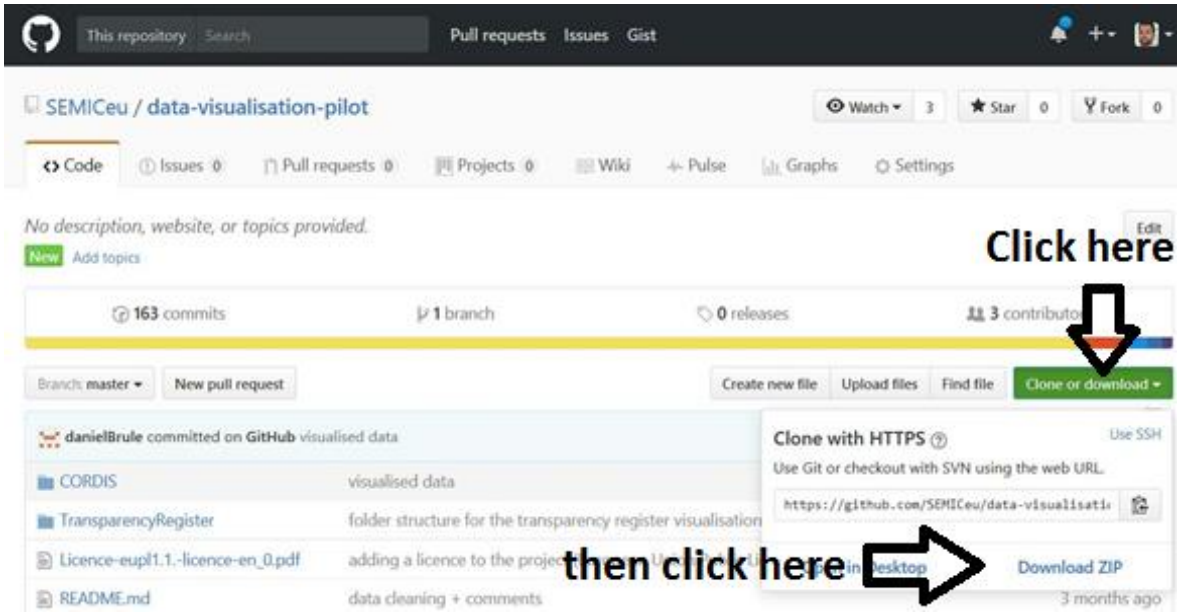


Figure 26: download of the GitHub project

I.3 Installation of Java SE Runtime Environment 8

The first step is to download a version 8 of Java SE Runtime Environment through the following link using a web browser of your choosing (e.g. Internet Explorer, Firefox, Google Chrome): http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html. Click on "Accept License Agreement" and download the file for your computer system, please refer to Figure 6 for a visual representation of the webpage.

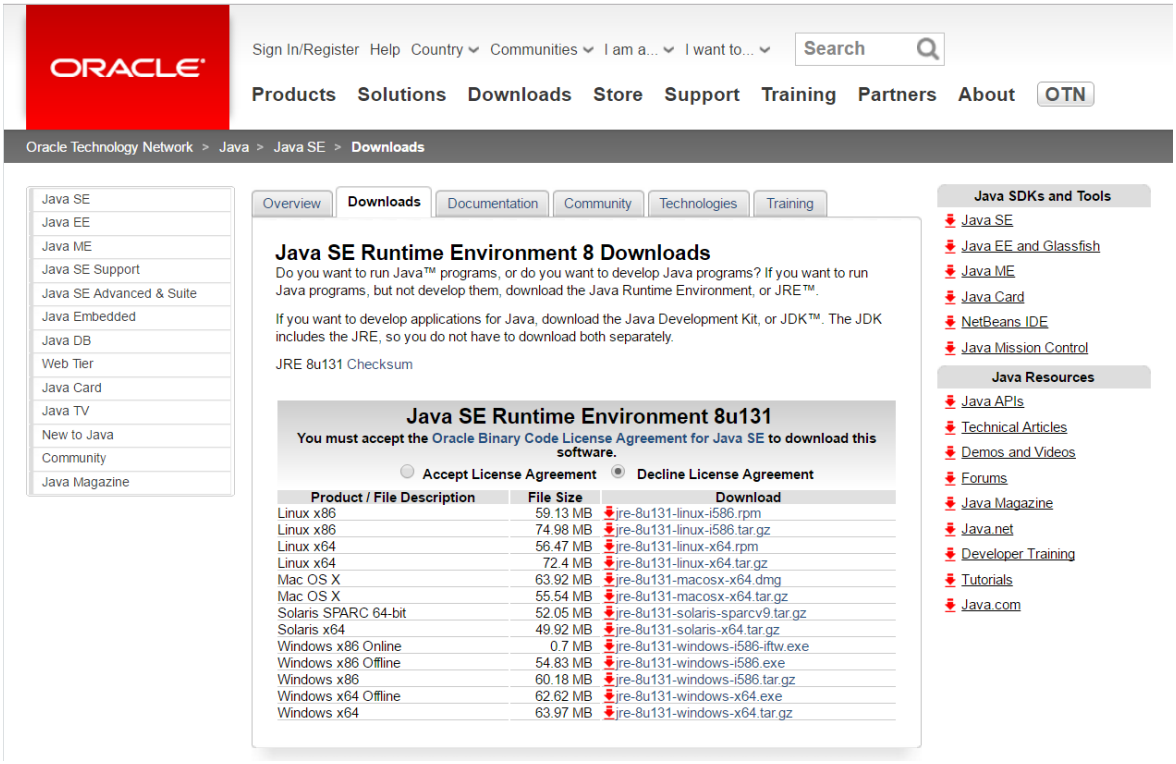
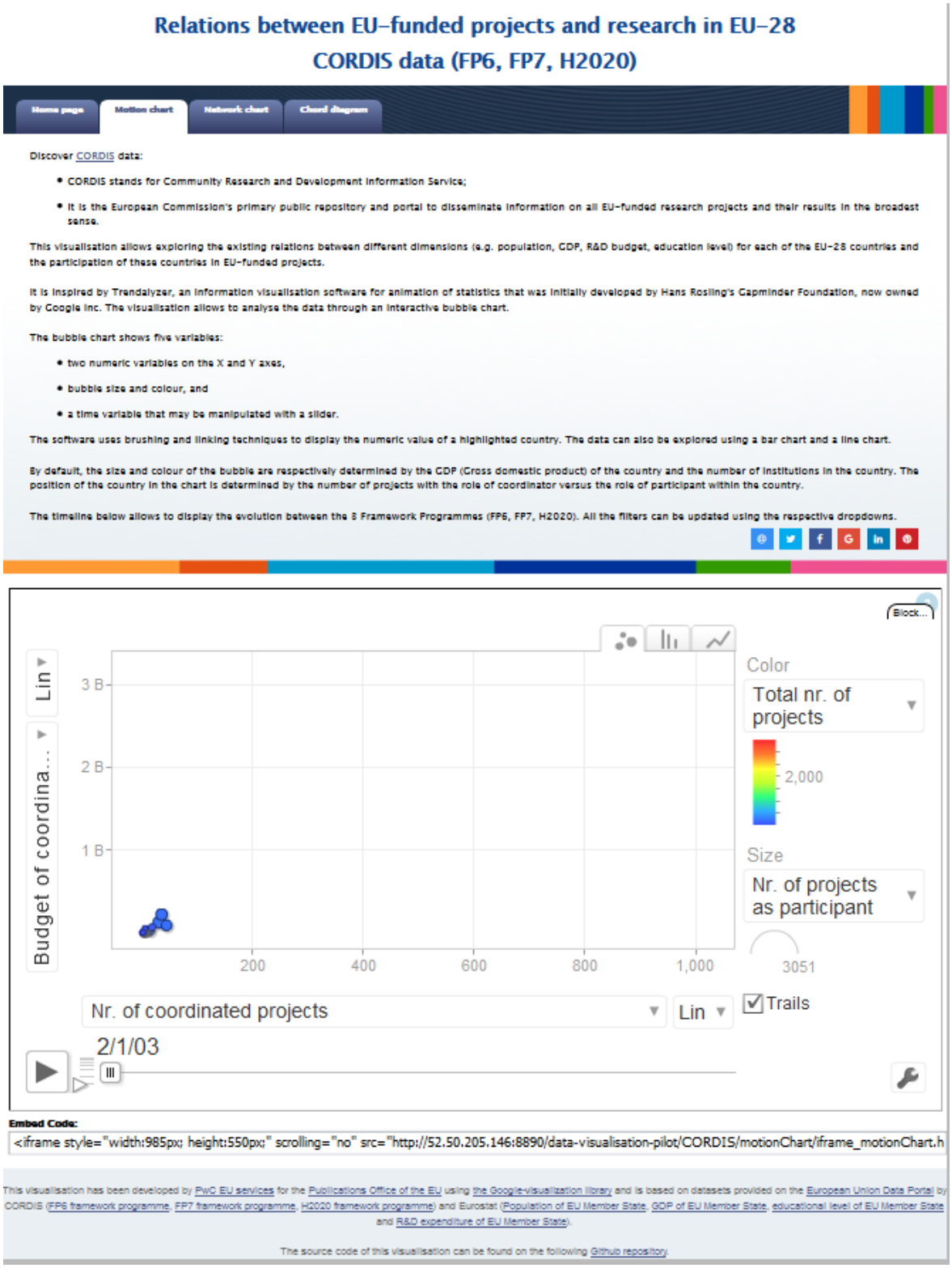


Figure 27 – Installation page Java SE Runtime Environment 8

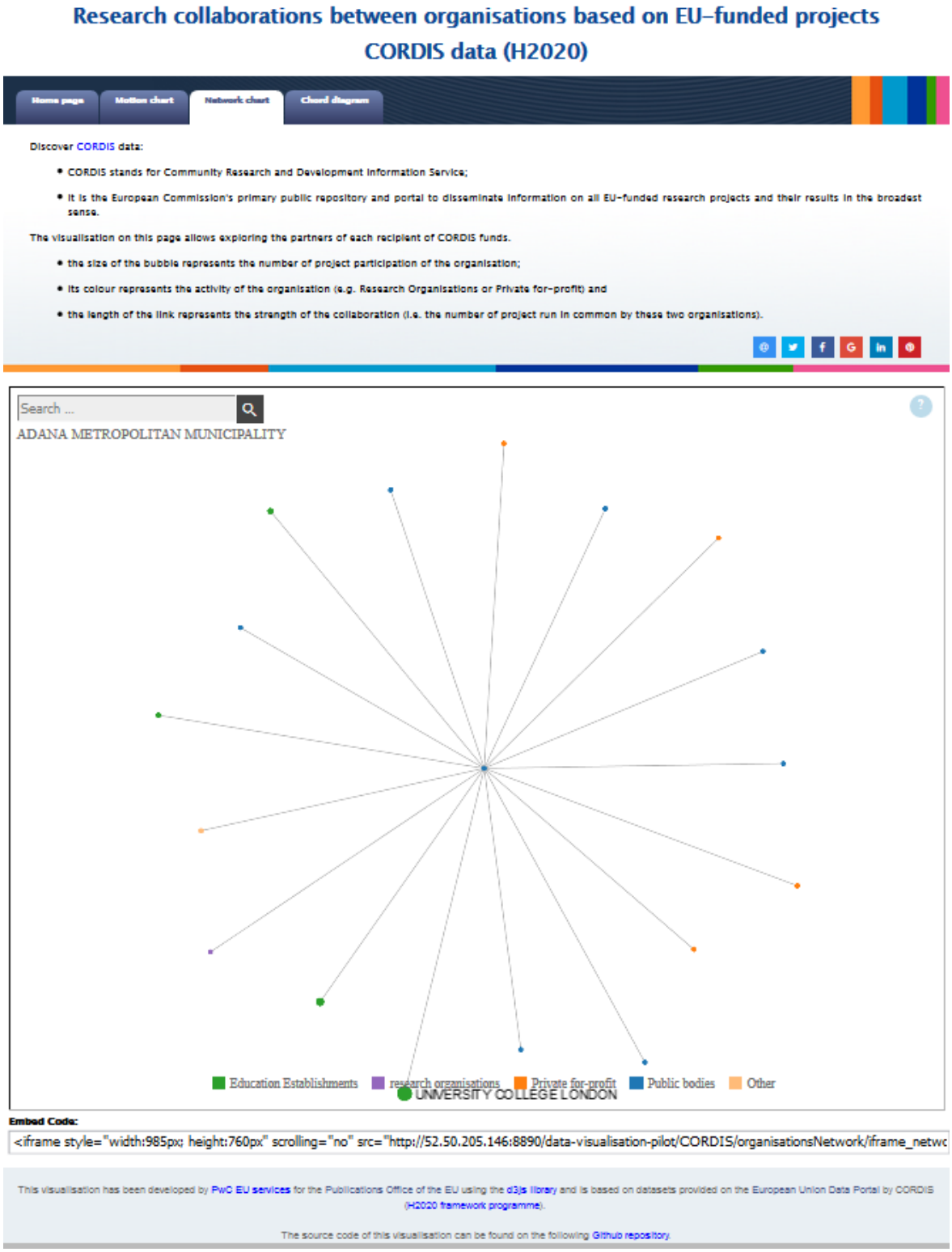
After downloading the installation file, go to the local folder on your computer where the installer has been downloaded (by default, the "Downloads" folder) and unzip the file.

Place the unzipped folder in C:\Program Files (x86)\Java on your computer.

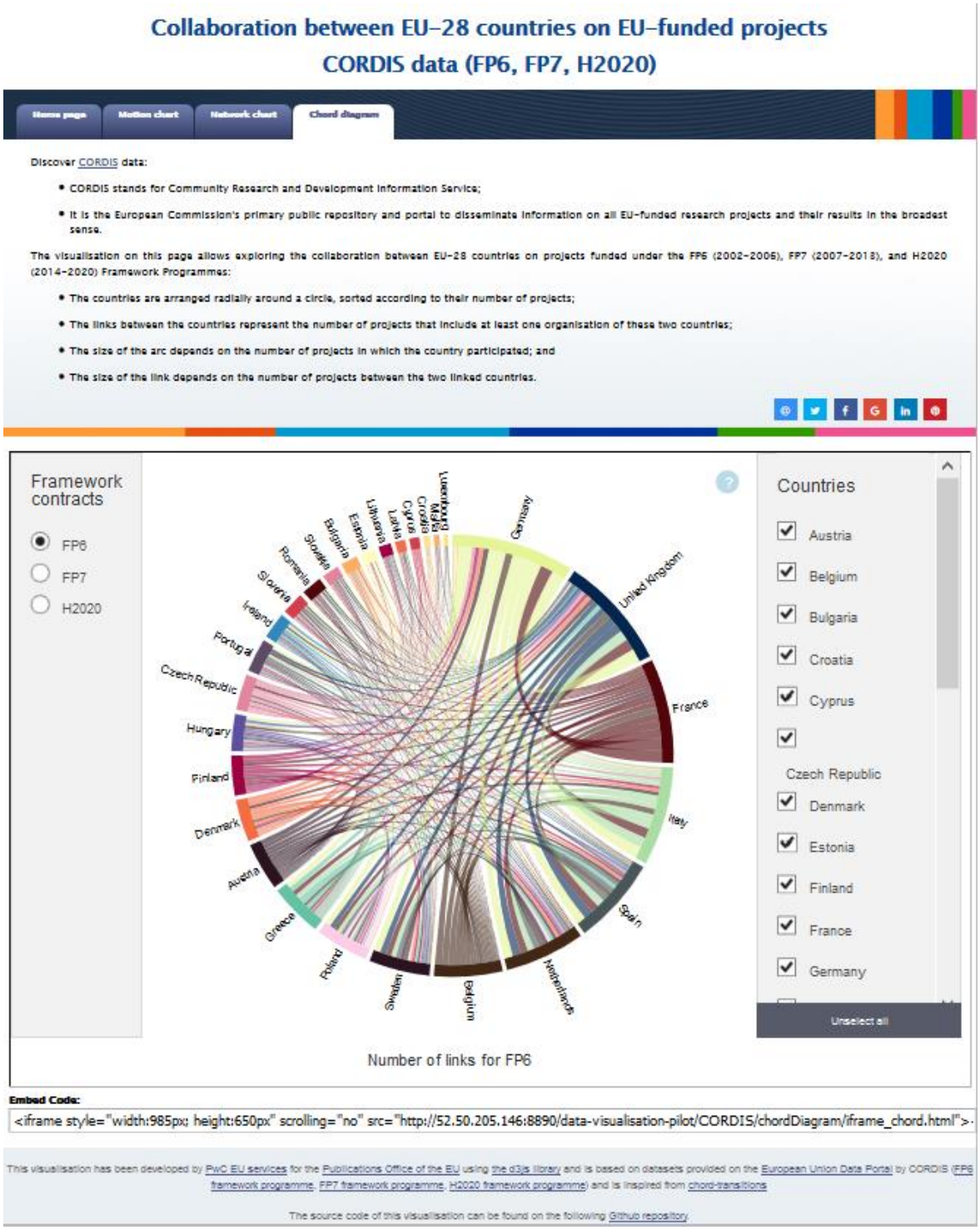
Annex IV. Motion chart web page



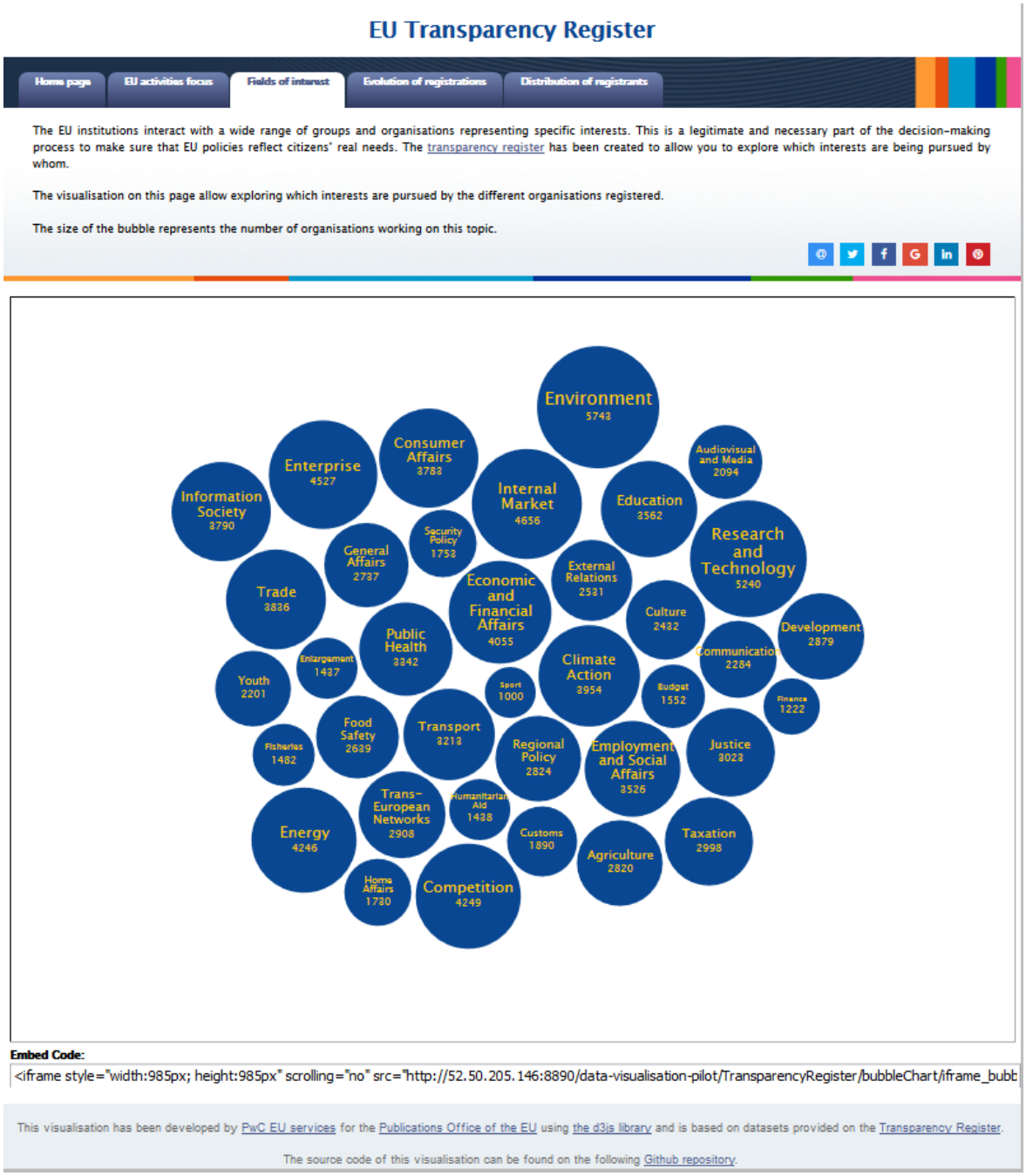
Annex V. Network diagram web page



Annex VI. Chord diagram web page



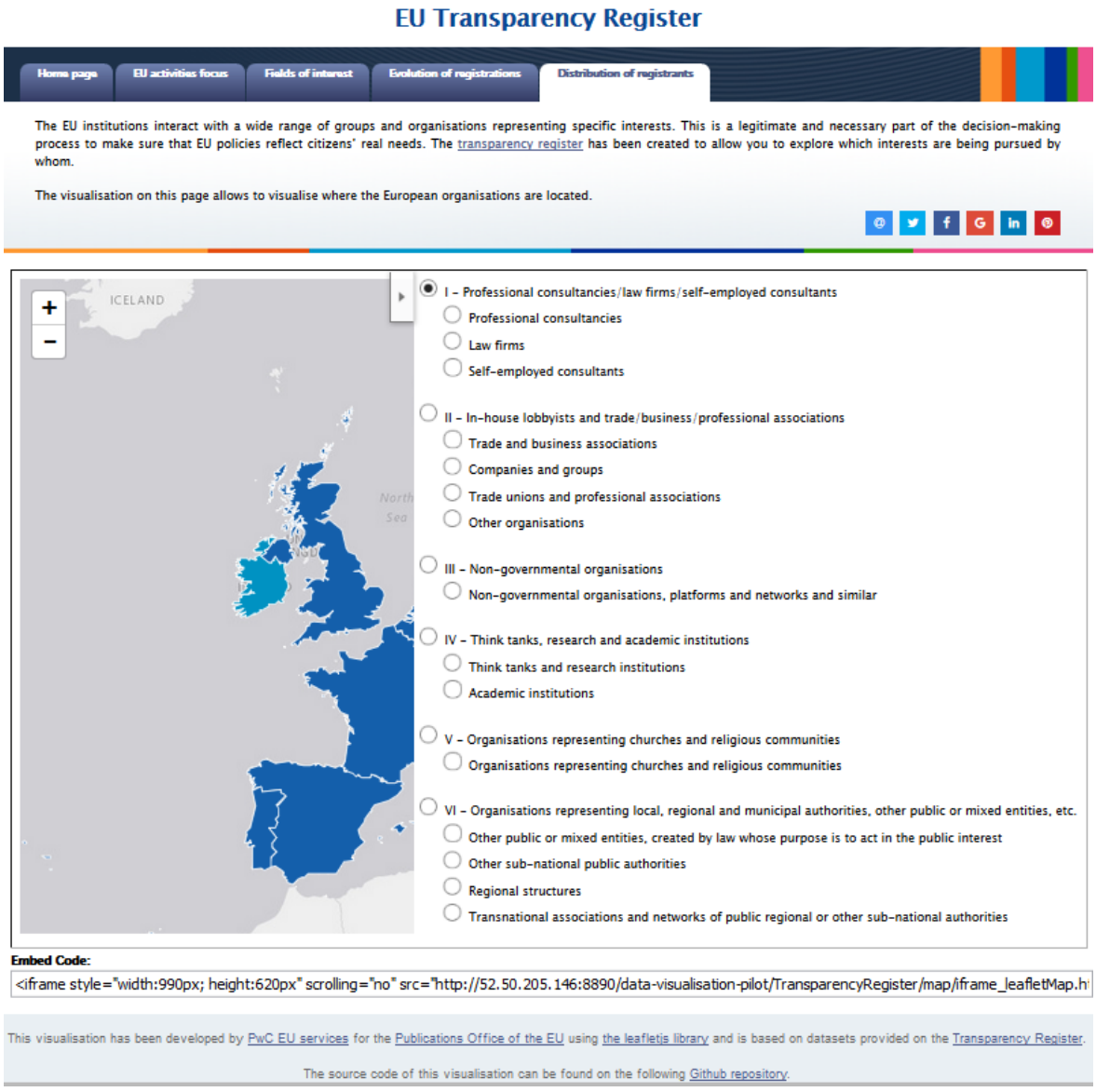
Annex VII. Bubble chart web page



Annex VIII. Wordcloud web page

Annex IX. Linechart web page

Annex X. Map web page



Annex XI. Global layout

A global layout, aligned on the one of the European Union Open Data Portal[73], has been created so every page contains a header with a title, tabs to navigate through every pages and a small separate section containing a description of the visualisation. Moreover, every page and visualisation is designed in such a way that it is compatible with low resolution screens (1360*768 px).

Another feature that is available on every page is the ability to share the page on the main social media. A set of icons (email, Twitter, Facebook, Google plus, LinkedIn and Pinterest) are placed under the description on the right to allow the visitor to share in one click, as it can be seen on Figure 28.

Figure 28: Global layout example

Finally, every visualisation is embedded in an iFrame. The code to include this iFrame on any website is provided below each visualisation. A simple copy-paste in any HTML page will allow to include the visualisation on another website.

[73] https://data.europa.eu/euodp/en/data