SC352DI07171

D02.04 Report on open data publishing and visualisation

Document Metadata

Date	2016-03-07
Status	internal
Version	0.03
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 and documents the visualisations realised in the context of "Task-02.4: Report on open data publishing and visualisation".

1.1 Context

The goal of the ISA² is to providing data publishers (institutions, agencies and other bodies) of the Open Data Portal and other stakeholders of the Open Data Portal (European Data Portal including affiliated MS portals) with an open data service package in order to enable them to further open up their data and increase data interoperability in view a better data reusability and data visualisation.

In the context of task-02.4 six sample visualisations of two high value datasets (CORDIS[1] and Transparency registered[2]) have been created. The data visualisations were built to look at the selected datasets from different perspectives. Each visualisation, although self-explanatory, is supported by a short text in the form of a data story which helps communicate its message to its target audience.

Commented [DB1]: To be updated

Commented [DB2]: To be updated

The selection of datasets have been done according to:

- 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 a home page (http://52.50.205.146:8890/data-visualisation-pilot/) hosted by PwC. This page gives 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)[3].

1.2 Structure and approach

This document is structured as follows:

Commented [NL3]: All this part needs to be rewritten. Please look at another deliverable from another project in order to see how we should start. What I would include in this section is what have we delivered, i.e. exactly which visualisations for which datasets, and also what was the process followed.

Commented [DB4R3]: Done

- Section 2 document 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 screenshot of the visualisations.
- Section 3 document 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 screenshot of the visualisations.
- Section 4 present a list of activities that has been realised in the context of this deliverables.

[1] https://data.europa.eu/euodp/en/data/dataset/cordisref-data
[2] https://data.europa.eu/euodp/en/data/dataset/transparency-register
[3] https://joinup.ec.europa.eu/community/eupl/og_page/european-union-public-licence-eupl-v11

2 CORDIS VISUALISATIONS

CORDIS stands for Community Research and Development Information Service. It 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[4];
- FP7: running from 2007 to 2013[5]; and
- H2020: running from 2014 to 2020[6].

2.1 Motion Chart using Gapminder

The visualisation (accessible at the following URL http://data-visualisation.semic.eu:8890/motionChart/motionChart.html) allows exploring the existing relations between different dimensions (e.g. population, GDP, R&D budget, education level) for each of the EU-28 countries and the participation of these countries in EU-funded projects.

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 are been extracted from CORDIS Datasets (framework contract FP6[7], FP7[8] and H2020[9]) and Eurostat (Population of EU Member State[10], GDP of EU Member State[11], population of educational attainment level[12] and Total intramural R&D expenditure[13]). The data used[14], the code developed for data preparation[15], the code used for the visualisation[16] as well as the data visualised[17] are available on GitHub.

Commented [NL5]: To be moved earlier – see my previous comment.
Then you can talk about the fields per dataset.

Commented [NL6R5]: I see later that you have a whole section on datasets used, so maybe you can move the links there and remove the duplication from here.

After reading later, I'm not sure if this introductory text is even needed. What do you think?

Commented [DB7R5]: I keep the introduction short. I keep the section dataset used in each section as each section used different CORDIS dataset

[4] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[5] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[6] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[7] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[8] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[9] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[10] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en
[11] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=naida_10_gdp&lang=en
[12] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_03&lang=en
[13] http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=rd_e_gerdtot&lang=en
[14] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets
[15] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/MotionChart/Data%20Preparation
[16] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/MotionChart/Visualisation
[17] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/MotionChart/VisualisedData

The visualisation allows to analyse the data through an interactive bubble chart (cf. Figure 1).

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.

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

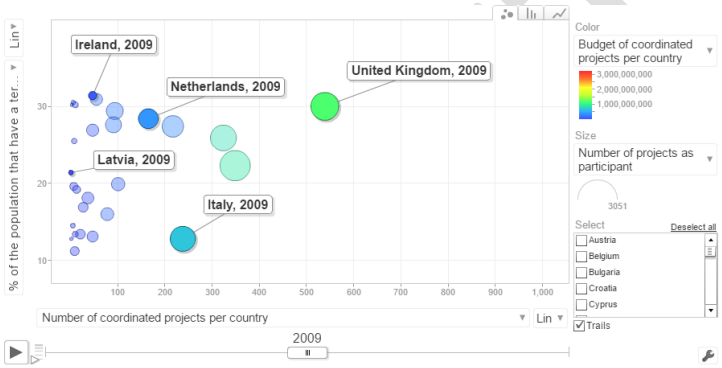


Figure 1: First screen of the motion chart - Bubble chart

The data can also be explored using a bar chart (cf. Figure 2) and a line chart (cf. Figure 3).

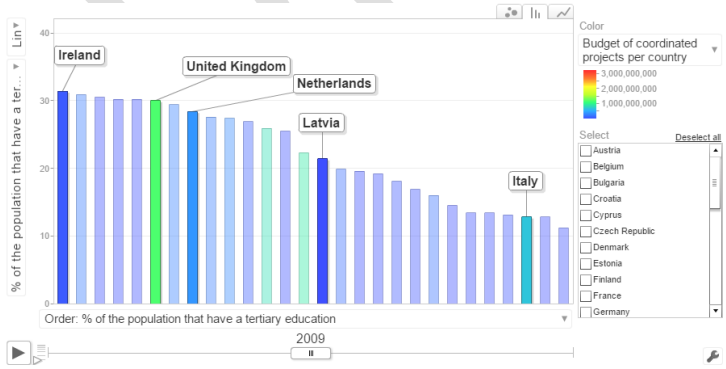


Figure 2: Second screen of the motion chart - Bar chart

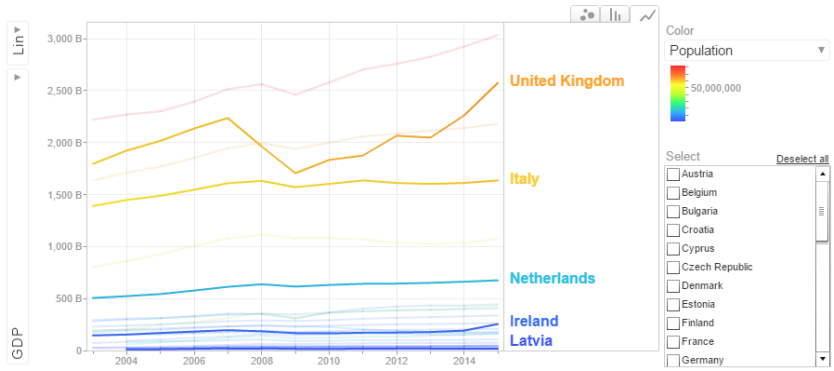


Figure 3: Third screen of the motion chart - Line chart

2.1.1 Activities

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

- An R script[18] was developed to gather and transform the data. The goal of this script was 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 was a CSV file[19] containing, for all countries, the dimensions listed in section 2.1 Motion Chart using Gapminder.
- To feed the bubble chart easily with data, the decision was made to use a Google spreadsheet[20] for hosting the data. This solution was the most convenient as it is user-friendly and directly supported by Google's visualisation library. The CSV file previously generated was imported in the Google document[21].
- Creation of a HTML code[22] that defines the structure of the web page.
- Creation of a JavaScript script (include in the HTML) using the Google-visualization library[23] to:
  - Query the Google spreadsheet 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

The HTML file containing the HTML, JavaScript and CSS code was uploaded on a PwC server (motionChart.html) and placed at the root folder. Moreover a link was added from the data-visualisation-pilot homepage that allows to access the chart by clicking on an image representing the bubble chart.

Commented [DB8]: Update with CSV

Commented [NL9]: Google should be written throughout with a capital G

Commented [DB10R9]: Done

Commented [NL11]: Which web page? You can refer to one of the figures above so that the reader understands what you're talking about.

Commented [DB12]: Add in annex a screen shot of the whole webpage

[18] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/MotionChart/Data%20Preparation/DataPrep-MotionChart.R

[19] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/Data%20cleaning%20Visualisation%201-V2/Output/CountryInformation.csv

[20] https://docs.google.com/spreadsheets/d/1Ex4EEEDeT8lfRWSi5nCwYnZBNskl9bMO-nmDX3mMszk/edit#gid=0

[21] https://gsuite.google.com

[22] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/MotionChart/Visualisation/motionChart.html

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

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.

2.1.4 Update of the visualisations

<TODO>

Commented [DB13]: To do after we know if we use R or python

- *Indicate which version of R/python is needed + link to annex III.1*
- *Download of source + if git available give command line to execute else link to annex III.2*
- *Indicate the command line to execute to update the data*

2.2 Network Diagram by organisation

The visualisation (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/organisationsNetwork/network_organisations_2.html) allows exploring the partners of each recipient of CORDIS funds.

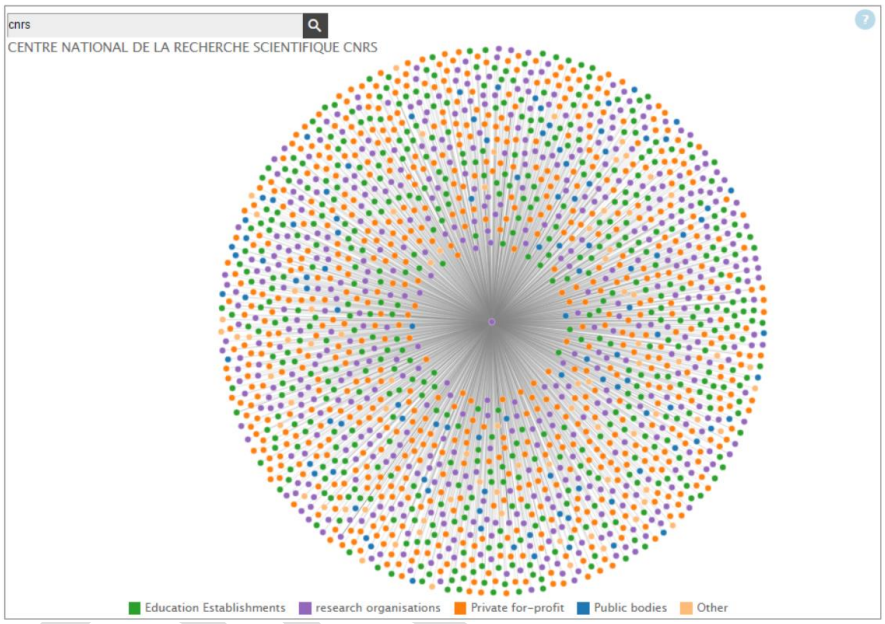


Figure 4: Network diagram per organisation

For one organisation:

- The size of the bubble represents the number of project participation of the organisation;
- Its colour represents the activity of the organisation (e.g. Research Organisations or Private for-profit) and
- The thickness of the link represents the strength of the collaboration (i.e. the number of project run in common by these 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[24]). The data used[25], the code developed for data preparation[26], the code used for the visualisation[27] as well as the data visualised[28] are available on GitHub.

[24] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects

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

[26] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/organisationsNetwork/Data%20Preparation

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

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

Only the project 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 describes the type of activity carried out by the organisation (cf. **Error! Reference source not found.**), 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[29] was developed to gather and transform the data. The goal of this script was, for each of the organisations, to create its network (i.e. organisation A work with B and C). Additional information (e.g. number of project realised by one organisation) was also added. The output of the script is a CSV file[30] containing, for all organisation, the needed information.
- A Python script[31] was created to read this CSV and transform it to JSON files formatted as needed to feed the network chart. One JSON file is available per organisation.
- The HTML code[32] defining the structure of the web page was created.
- The JavaScript[33] code was adapted from the D3 example[34] to:
  - Read the JSON 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 (network_organisations_2.html, network_organisations.js, orgs_filter.js, network-org-style.css, filtering_panel.css). Furthermore, the JSON files containing the data were placed in the "Data" subfolder as well as the organisations.js JavaScript file containing the list of organisations to feed the filtering panel. 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 has been realised in the context of this visualisation.

2.2.4 Update of the visualisations

<TODO>

Commented [DB14]: To do after we know if we use R or python

- *Indicate which version of R/python is needed + link to annex III.1*
- *Download of source + if git available give command line to execute else link to annex III.2*

[29] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/organisationsNetwork/Data%20Preparation/DataPrep-Orglink.R

[30] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/organisationsNetwork/VisualisedData/NbLinkOrgLevel1.csv

[31] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/organisationsNetwork/Data%20Preparation/GenerateJSONforOrg_multipleOrgs.py

[32] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/organisationsNetwork/Visualisation/network_organisations_2.html

[33] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/organisationsNetwork/Visualisation/network_organisations.js

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

- *Indicate the command lines to execute to update the data*

2.3 Chord Diagram

The visualisation (accessible at the following URL http://52.50.205.146:8890/data-visualisation-pilot/chord-diagram/chart/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.

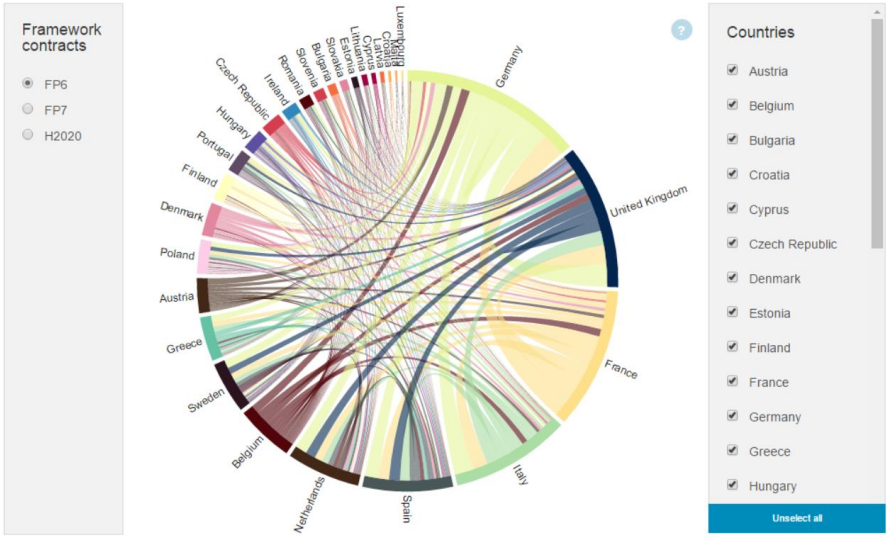


Figure 5: Chord diagram

The data visualised are been extracted from CORDIS Datasets (framework contract FP6[35], FP7[36] and H2020[37]). The data used[38], the code developed for data preparation[39],

[35] https://data.europa.eu/euodp/data/dataset/CORDISfp6projects
[36] https://data.europa.eu/euodp/data/dataset/CORDISfp7projects
[37] https://data.europa.eu/euodp/data/dataset/CORDISH2020projects
[38] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/Datasets
[39] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/chord-diagram/Data%20Preparation/DataPrep-Link.R

the code used for the visualisation[40] as well as the data visualised[41] are available 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[42] 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 organisation 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[43] that defines the structure of the web page.
- Creation of a JavaScript script using the Google-visualization library[44] to:
  - Read the data;
  - Configure the chart;
  - 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 chord-diagram containing sub-folders. The "data" subfolder contains the nbLinks.csv file which contains the data used to draw the chord diagram. 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.

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

2.3.3 Configuration done

The bigger 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 one to highlight/grey out countries.

2.3.4 Update of the visualisations

<TODO>

Commented [DB15]: To do after we know if we use R or python

- *Indicate which version of R/python is needed + link to annex III.1*
- *Download of source + if git available give command line to execute else link to annex III.2*

[40] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chord-diagram/Visualisation

[41] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chord-diagram/VisualisedData

[42] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/chord-diagram/Data%20Preparation/DataPrep-Link.R

[43] https://github.com/SEMICeu/data-visualisation-pilot/blob/master/CORDIS/chord-diagram/Visualisation/chart/chord.html

[44] https://github.com/SEMICeu/data-visualisation-pilot/tree/master/CORDIS/chord-diagram/Visualisation/js

- *Indicate the command lines to execute to update the data*

3 OTHER DATASET

Commented [DB16]: To be complete with the additional dataset

4 ACTIVITIES

- 2017-01-12: Meeting with CORDIS team.
- 2016-12: Work on the various visualisations requested by OP: Gapminder, 2 Chord visualisations, network of organisations, and network diagram.
- 2016-12-02: data preparation for the Chord and Network visualisation (country and organisation)
- 2016-11-28: Submission of updated visualisations of CORDIS (types: Bubble charts & Network) http://data-visualisation.semic.eu:8890/
- 2016-11-28: Request from OP for another visualisation on CORDIS data in a network diagram, visualising the data by organisation.
- 2016-11-25: Submission of CORDIS data visualisation Gapminder.pptx http://data-visualisation.semic.eu:8890/gapminder.html
- 2016-11-24: Submission of CORDIS visualisation draft http://data-visualisation.semic.eu:8890/
- 2016-11-18: Call to discuss CORDIS visualisation requirements. Request from OP for 3 visualisations:

1) Bubble (Gapminder) on
   - GDP and Population of the countries
   - # of project participation
   - # of project coordinated
   - # of institution per country
2) Network Diagram
   - We will be able to select, through a dropdown, if we work at country or institution level
   - Dots: Country or institution
   - Lines: number of project between the two countries/institutions
   - Colour: if institution is selected then there will be one colour per country
     if country is selected then there will be one colour for all the dots
   - Interactions: if a link/dot is selected it will highlight
3) Same as 2) but using a chord diagram.

- 2016-11-14/18: Analysis and preparation of a proposal for CORDIS visualisations based on OP feedback (10/11).
- 2016-11-15: ES installed visualization tool used by India Open Data tool: the advice is not to use the tool as it is not mature, the development stopped, and the code available doesn't reflect the one online
- 2016-10-27: Submission of CORDIS data visualisation proposal and data story, integrating AZ's comments.
- 2016-10-26: Submission of CORDIS data visualisation proposal and data story draft.
- 2016-10-24: AZ sent the references to take into account for the task on CORDIS data publishing and visualisation:
  - An article on the Spiegel on analysing different networks: http://www.spiegel.de/wissenschaft/mensch/horizon-2020-eu-forschungsgelder-sind-ungleich-verteilt-a-1053177.html
  - Spaghetti-open-data/eu research funding network is reusing CORDIS data from ODP on FP7 and Horizon 2020. The aim is to compare the networks of partnership in both programmes. Spaghetti community aims to produce recommendation for the Commission on how to release this type of data in the future to enable reuse and civic hacking.
- 2016-10-24: Kick-off call. Alignment on scope and next steps.

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 found.**)
Address (including country)	Address of the organisation

Annex III. How to update the Motion Chart (Gapminder) data

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.

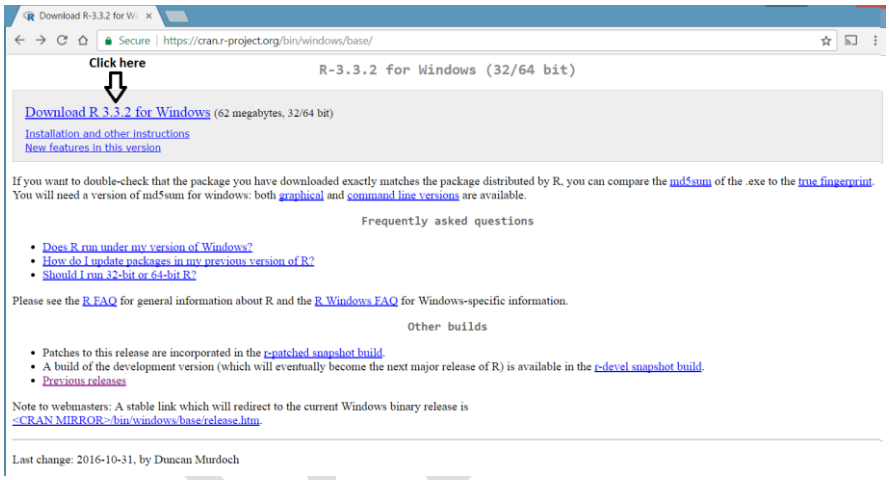


Figure 6: installation page R

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.

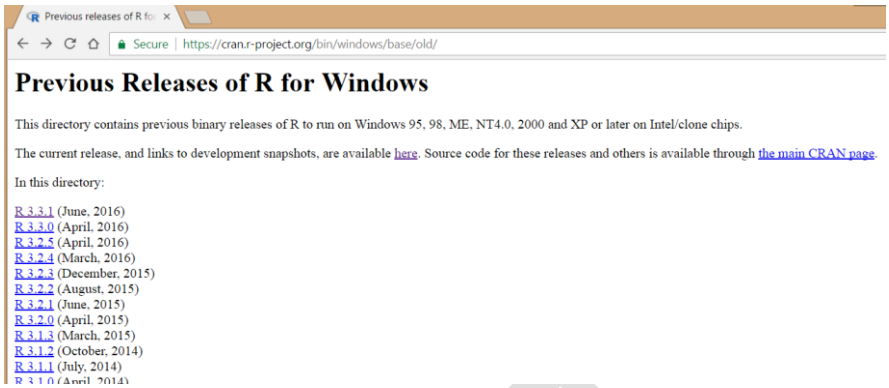


Figure 7: previous releases of R

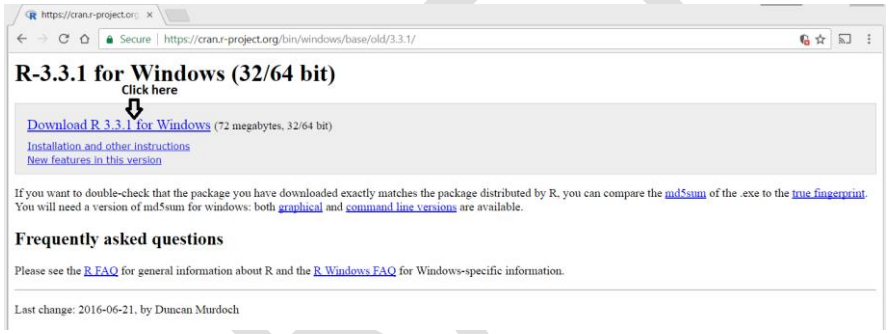


Figure 8: installation page of previous releases of R

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".

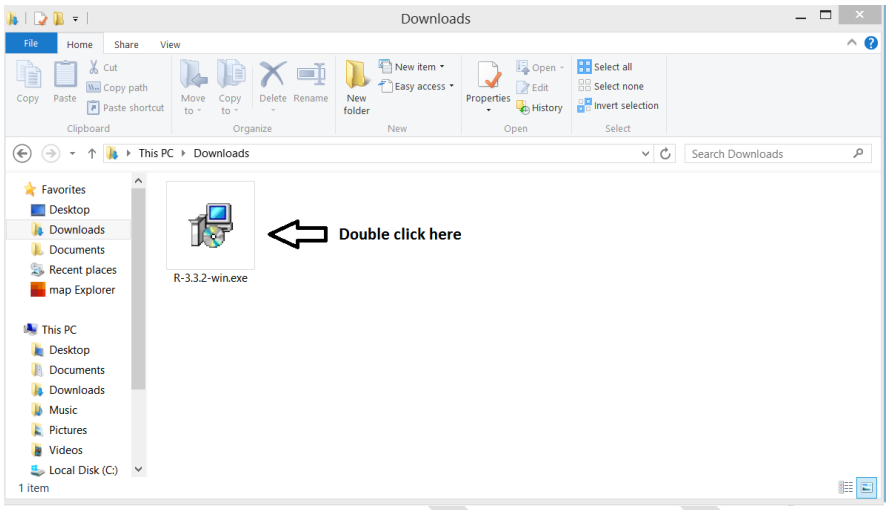


Figure 9: opening the downloaded installation file for R

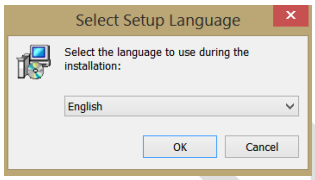


Figure 10: selection of installation language

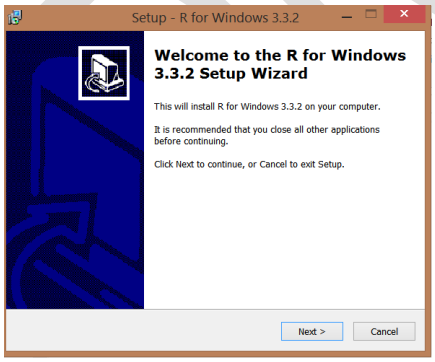


Figure 11: installation screen R (1)

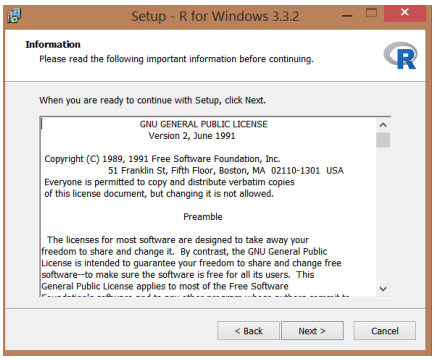


Figure 12: installation screen R (2)

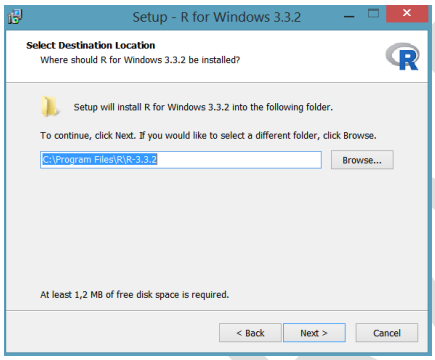


Figure 13: installation screen R (3)

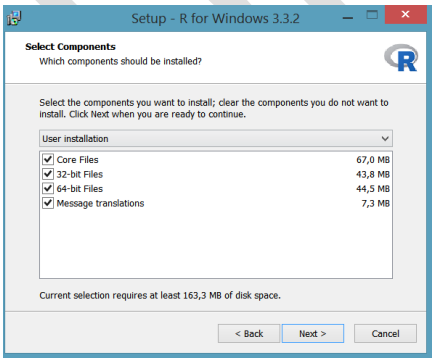


Figure 14: installation screen R (4)

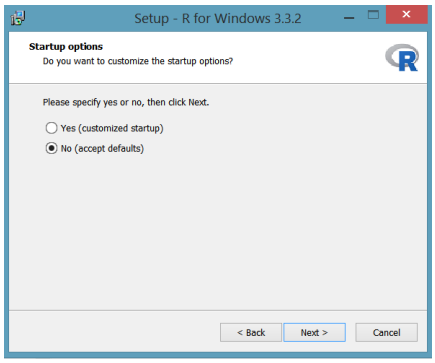


Figure 15: installation screen R (5)

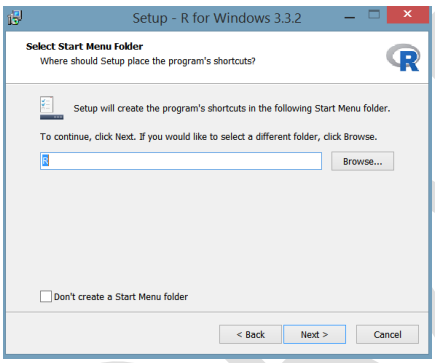


Figure 16: installation screen R (6)

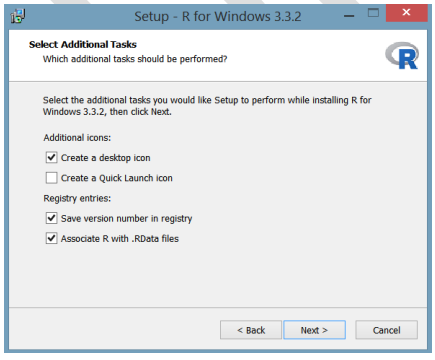


Figure 17: installation screen R (7)

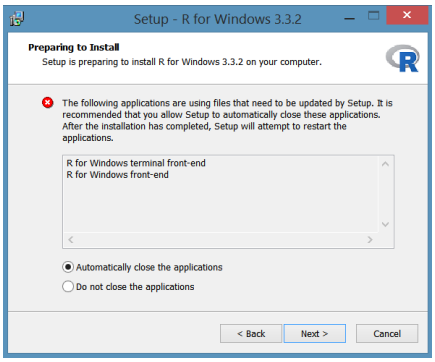


Figure 18: installation screen R (8)

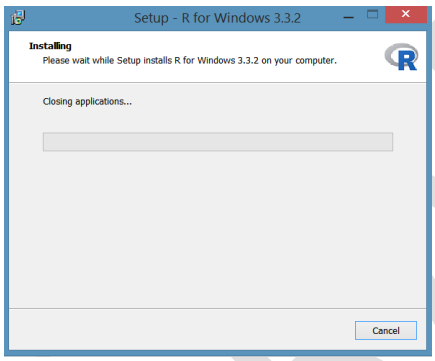


Figure 19: installation screen R (9)

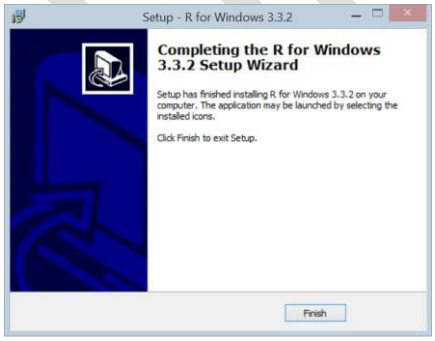


Figure 20: installation screen R (10)

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.

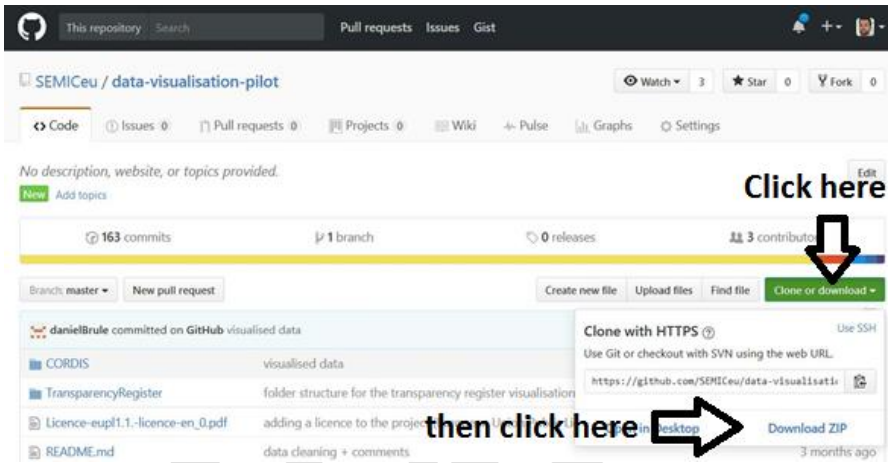


Figure 21: download of the GitHub project

I.3 Downloading the most recent version of the data required

I.3.1 CORDIS data (FP6, FP7, H2020)

Go to https://data.europa.eu/euodp/data/dataset/cordisfp6projects, download the "FP6 organisations" csv file and the "FP6 projects" csv file. Please refer to Figure 22 for a visual representation of this. Repeat the same process for FP7 (https://data.europa.eu/euodp/data/dataset/cordisfp7projects) and H2020 (https://data.europa.eu/euodp/data/dataset/cordisH2020projects).

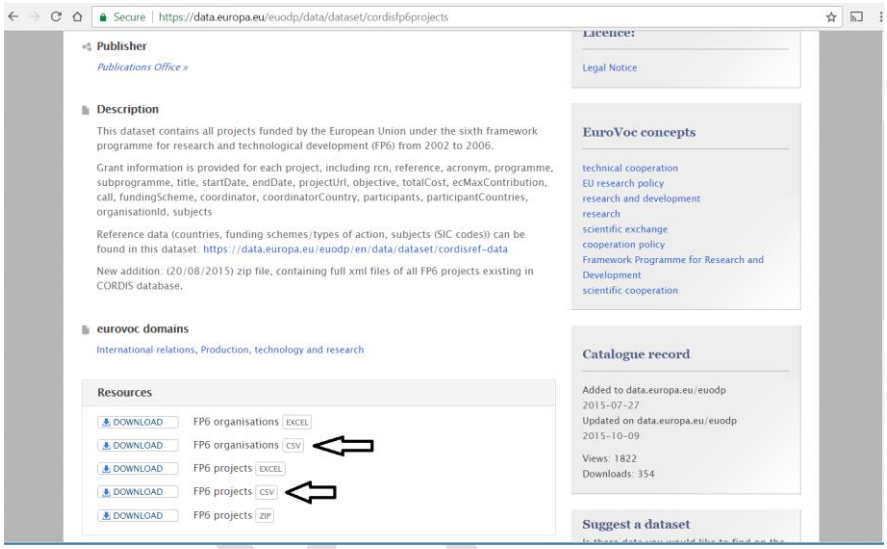


Figure 22: downloading CORDIS data

I.3.2 Eurostat data (Population, GDP, population by education level, and total intramural R&D expenditure)

Go to http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en : demo_pjan_1_Data.csv, refer to Figure 23 for a visual representation of the page.

Next, click on the "+" icon behind GEO, as indicated in Figure 23. A window will appear as in Figure 24. In this window, tick all EU28 countries and untick all others.

Commented [CV17]: @Daniel, should she select DE or DE_TOT for Germany?

Afterwards, go to the "TIME" tab of the pop-up window and select from 2013 to the current year. Make sure all others are deselected. Please refer to Figure 25.

Click "Update", the pop-up window will now close and you will return to the view as in Figure 23, updated with your selected filters.

Click on "Download" on the right top of the screen (see Figure 26). A new screen will appear. Here, make sure that "Full extraction (1 data tables)", "Single file" and "1 234.56" in the CSV part of the screen are selected, the rest should be deselected (Please refer to Figure 27). Finally, click "Download in CSV Format".

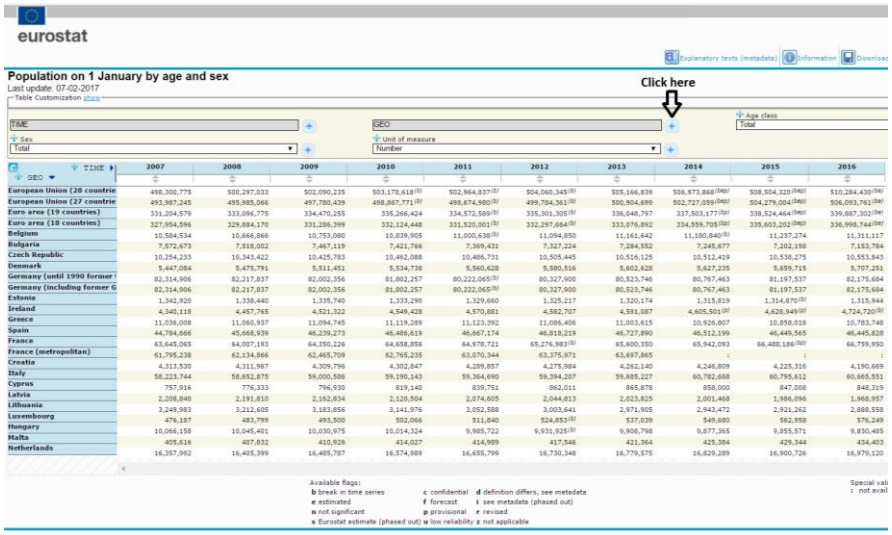


Figure 23: Eurostat - population database

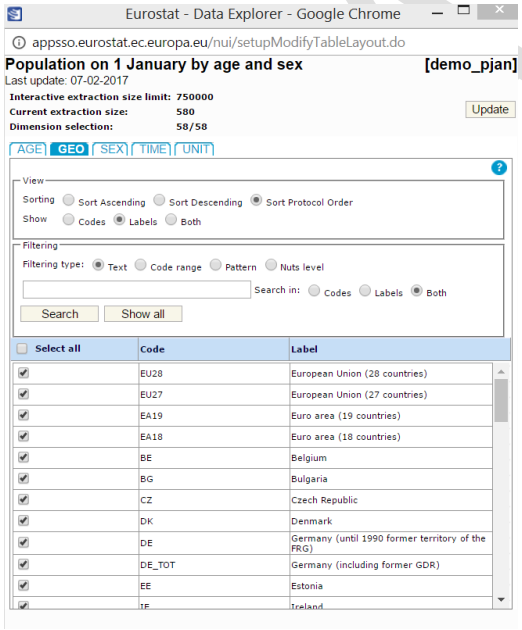


Figure 24: Eurostat - GEO filter

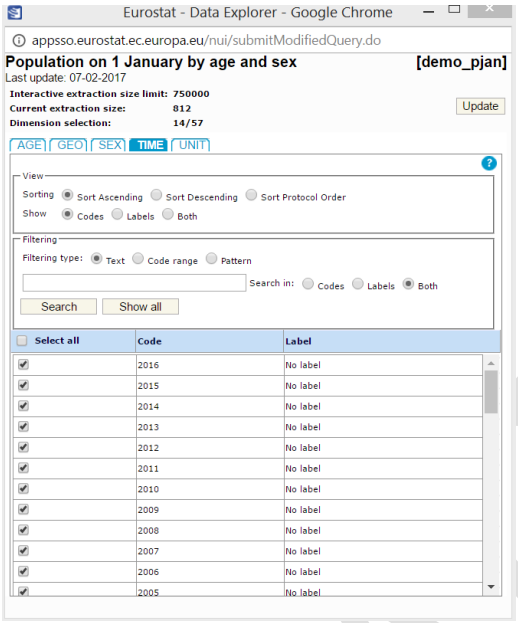


Figure 25: Eurostat - TIME filter

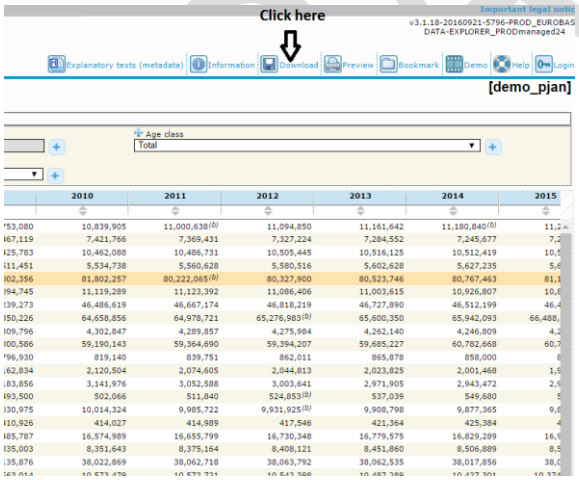


Figure 26: Eurostat - go to download

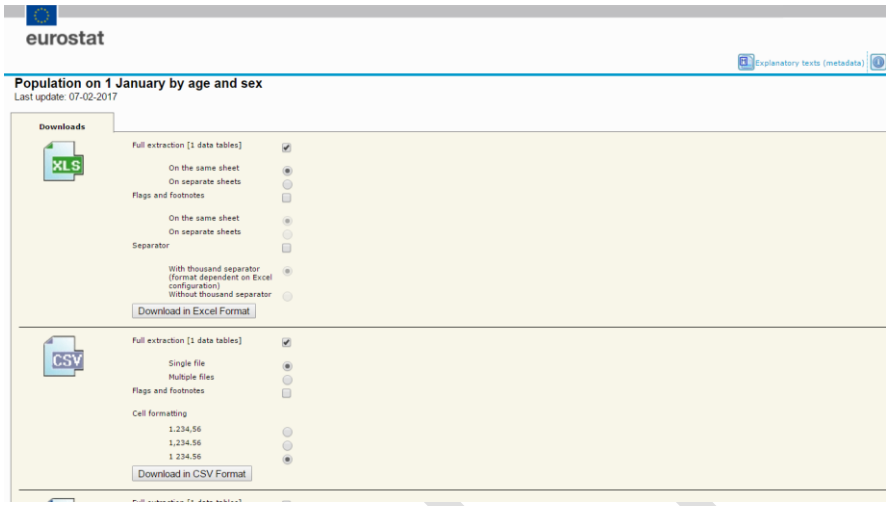


Figure 27: Eurostat - download CSV

Follow the procedure as described above to download the data on "Population by educational attainment level (%) - main indicators". You can reach this data on the following link:

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=edat_lfse_03&lang=en_-_edat_lfse_03_1_Data.csv.

Follow the same procedure to download the GDP data through the following link:

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=naida_10_gdp&lang=en_:_naida_10_gdp_1_Data.csv.

Make sure to apply the "UNIT" filter "Current prices, million euro" and the "NA_ITEM" filter "Gross domestic product at market prices" after having selected the "GEO" and "TIME" filter (Figure 24 and 25), but before clicking "Update". Please refer to Figure 28 and 29 to see a visual representation of the "UNIT" and "NA_ITEM" filter respectively.

To Download the "Total intramural R&D expenditure (GERD) by sectors of performance", go to the following link:

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=rd_e_gerdtot&lang=en_:_rd_e_gerdtot_1_Data.csv.

Make sure to apply the "UNIT filter "Million euro" after having selected the "GEO" and "TIME" filter (Figure 24 and 25), but before clicking "Update". Please refer to Figure 30 to see a visual representation of the "UNIT" filter.

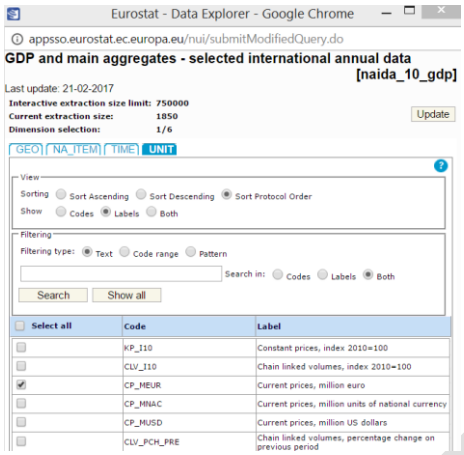


Figure 28: Eurostat - UNIT filter for GDP

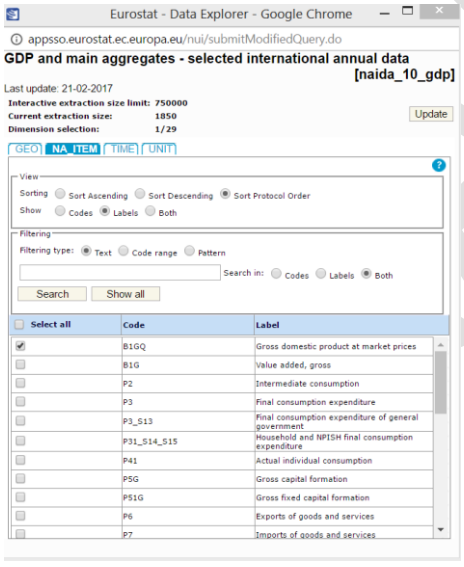


Figure 29: Eurostat - NA_ITEM filter for GDP

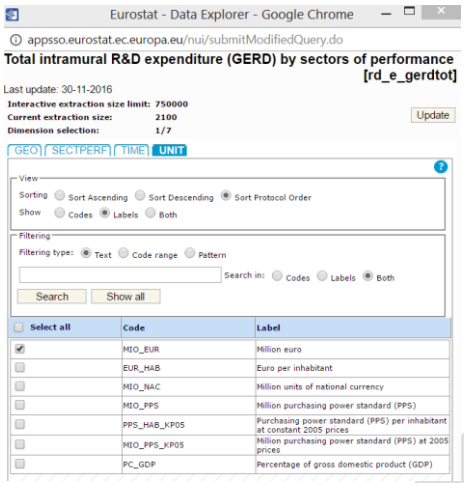


Figure 30: Eurostat - UNIT filter for R&D expenditure

After downloading all files, place all of them in the "Datasets" folder of the GitHub project (cf. the folder chosen by you when unzipping the downloaded GitHub project in step I.2) on your drive replacing the existing files. Refer to Figure 31 for a visual representation of this.

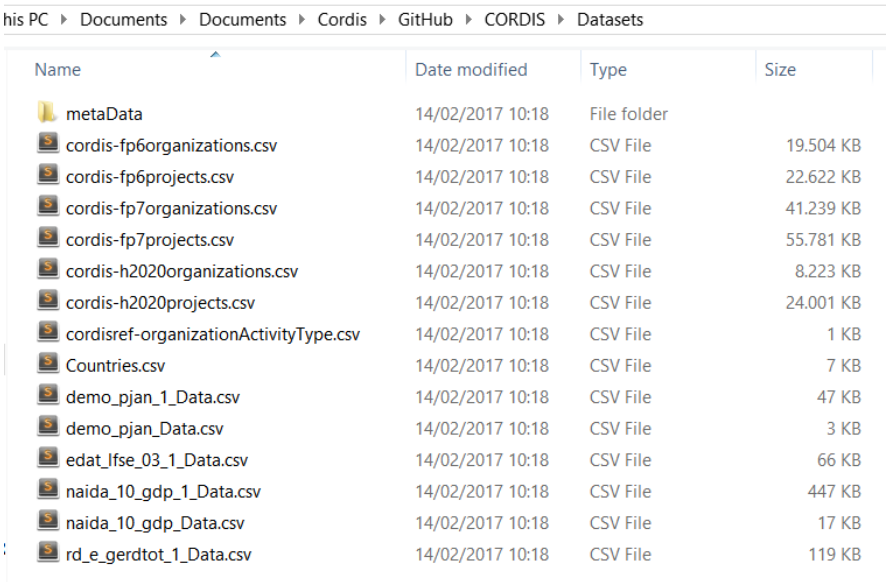


Figure 31: Datasets folder of the GitHub project

I.4 Launching the R script

Press "Windows + R" to open the "Run" box, here type "cmd" and click "OK" (see Figure 32).

The Command Prompt will open (see Figure 33).

Type "cd" followed by a space followed by the path to the location on your computer were you placed the unzipped GitHub project (see step I.2) and press enter (see Figure 34).

Type "cd" followed by a space followed by "CORDIS\MotionChart" and press enter (see Figure 35).

Type "C:\Program Files\R\R-3.3.2\bin\Rscript.exe" followed by a space followed by "Data Preparation\ DataPrep-MotionChart.R" followed by a space followed by the current year and press enter. Wait a moment for the script to finish, this can take up to a few minutes (see Figure 36).

Finally, type "exit" and press enter, the Command Prompt will close.

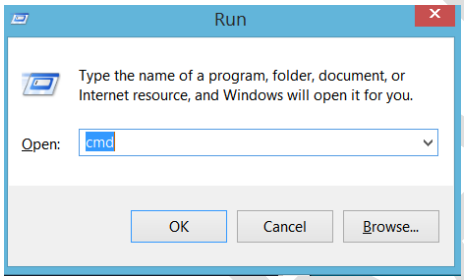


Figure 32: "Run" box

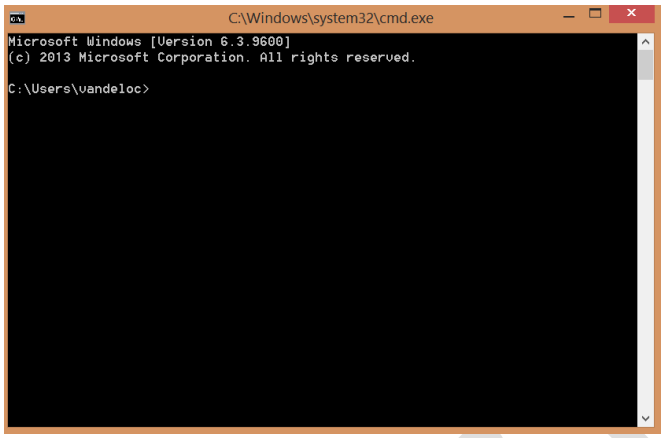


Figure 33: Command Prompt (1)

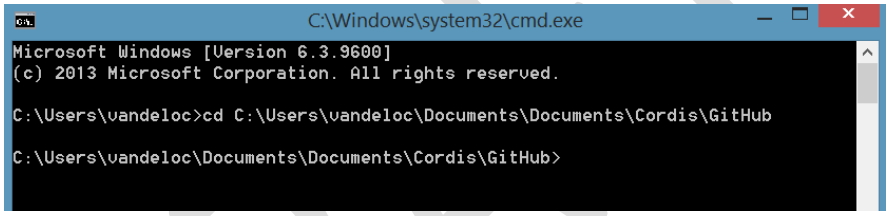


Figure 34: Command Prompt (2)

Figure 35: Command Prompt (3)

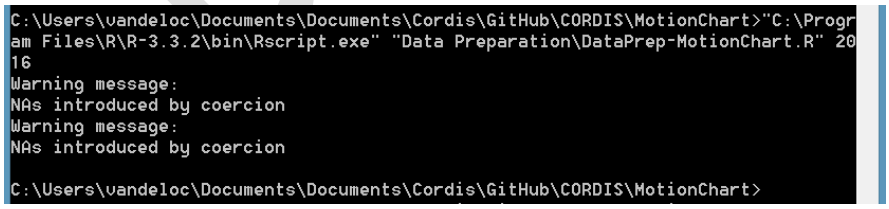


Figure 36: Command Prompt (4)

I.5 Copying into the Google spreadsheet

Go to the location on your computer were you placed the unzipped GitHub project (see step I.2) and go to the subfolder "CORDIS\MotionChart\VisualisedData". Right-click the data file in this location and select "Open with" then select "Excel" in the sub-menu (see Figure 37).

In Excel, make sure the entire "A" column is selected by clicking on the "A" on top of it and go to the "Data" tab and select "Text to Columns" (see Figure 38).

In the "Text to Columns windows, click "next" and finally "Finish. Please refer to Figure 30 to 41, make sure "Tab" is selected in Figure 40.

Delete the title row (see Figure 42), select all remaining content and copy (Ctrl + C).

Go to the Google spreadsheet on the following link:

https://docs.Google.com/spreadsheets/d/1Ex4EEEDeT8lfRWSi5nCwYnZBNskl9bMO-nmDX3mMszk/edit?ts=58385836#gid=0.

Select the B2 cell and paste (Ctrl + V) (see Figure 43). Scroll down to the bottom of the file and make sure the "A" column is completed (an ID going up one number each time) until the last row with data.

The visualisation is now updated with the most recent data.

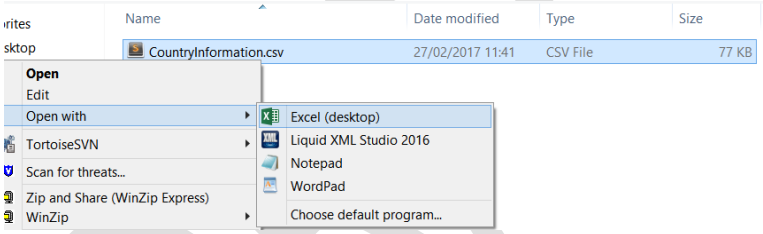


Figure 37: open file in Excel

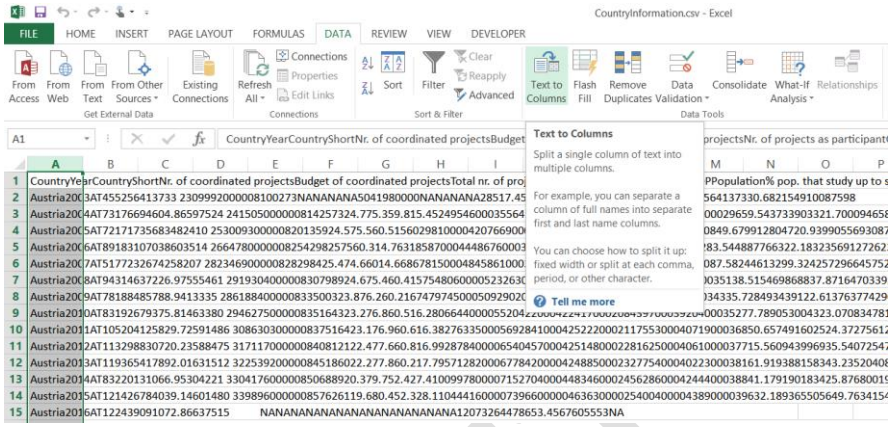


Figure 38: open "Text to Columns"

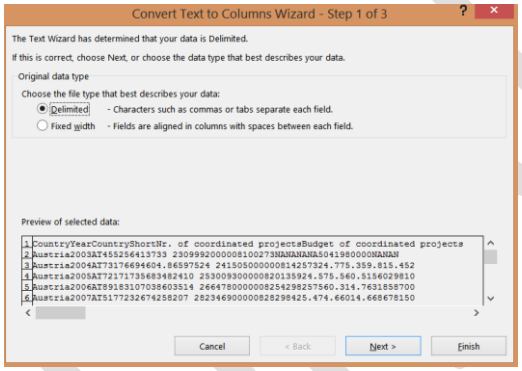


Figure 39: Text to Columns (1)

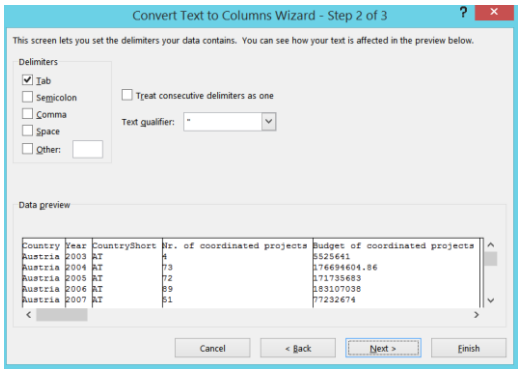


Figure 40: Text to Columns (2)

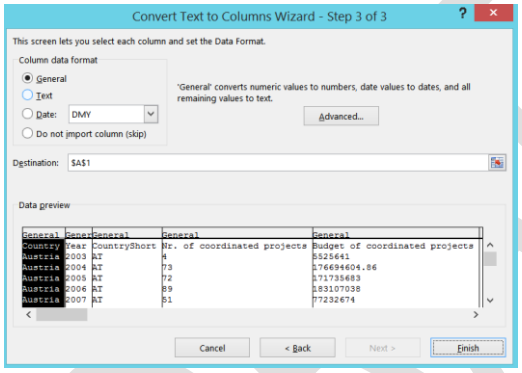


Figure 41: Text to Columns (3)

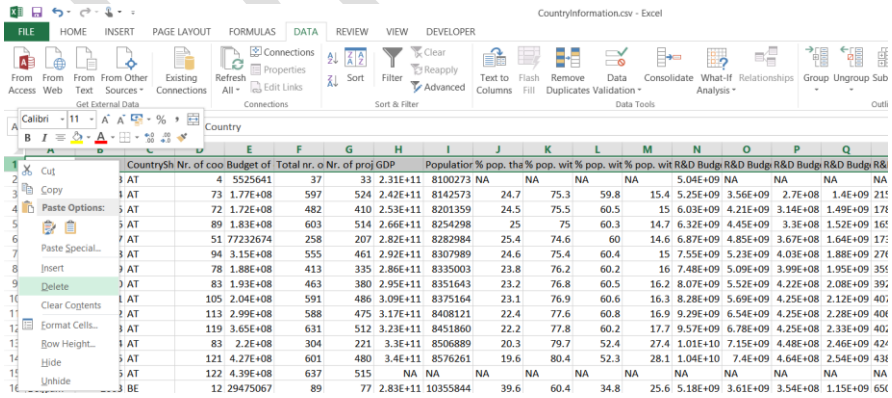
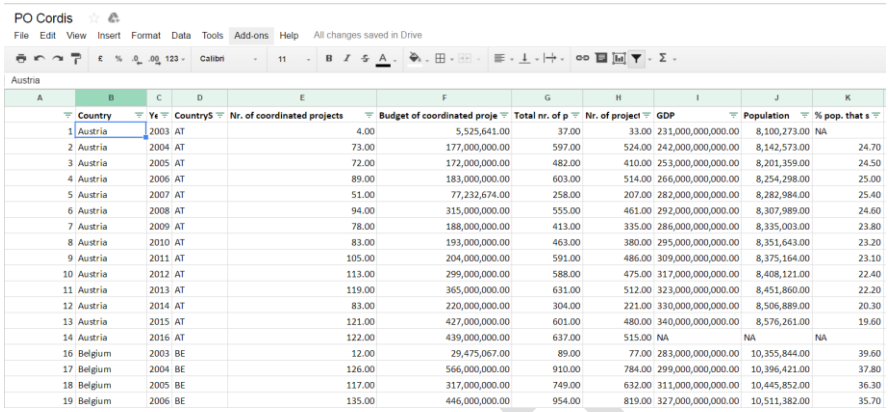


Figure 42: delete title row



PO Cordis

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Calibri 11

Austria

	A	B	C	D	E	F	G	H	I	J	K
		Country	Ye	CountryS	Nr. of coordinated projects	Budget of coordinated proje	Total nr. of p	Nr. of project	GDP	Population	% pop. that s
1		Austria	2003	AT	4.00	5,525,641.00	37.00	33.00	231,000,000,000.00	8,100,273.00	NA
2		Austria	2004	AT	73.00	177,000,000.00	597.00	524.00	242,000,000,000.00	8,142,573.00	24.70
3		Austria	2005	AT	72.00	172,000,000.00	482.00	410.00	253,000,000,000.00	8,201,359.00	24.50
4		Austria	2006	AT	89.00	183,000,000.00	603.00	514.00	266,000,000,000.00	8,254,298.00	25.00
5		Austria	2007	AT	51.00	77,232,674.00	258.00	207.00	282,000,000,000.00	8,282,984.00	25.40
6		Austria	2008	AT	94.00	315,000,000.00	555.00	461.00	292,000,000,000.00	8,307,989.00	24.60
7		Austria	2009	AT	78.00	188,000,000.00	413.00	335.00	286,000,000,000.00	8,335,003.00	23.80
8		Austria	2010	AT	83.00	193,000,000.00	463.00	380.00	295,000,000,000.00	8,351,643.00	23.20
9		Austria	2011	AT	105.00	204,000,000.00	591.00	486.00	309,000,000,000.00	8,375,164.00	23.10
10		Austria	2012	AT	113.00	299,000,000.00	588.00	475.00	317,000,000,000.00	8,408,121.00	22.40
11		Austria	2013	AT	119.00	365,000,000.00	631.00	512.00	323,000,000,000.00	8,451,860.00	22.20
12		Austria	2014	AT	83.00	220,000,000.00	304.00	221.00	330,000,000,000.00	8,506,889.00	20.30
13		Austria	2015	AT	121.00	427,000,000.00	601.00	480.00	340,000,000,000.00	8,576,261.00	19.60
14		Austria	2016	AT	122.00	439,000,000.00	637.00	515.00	NA	NA	NA
16		Belgium	2003	BE	12.00	29,475,067.00	89.00	77.00	283,000,000,000.00	10,355,844.00	39.60
17		Belgium	2004	BE	126.00	566,000,000.00	910.00	784.00	299,000,000,000.00	10,396,421.00	37.80
18		Belgium	2005	BE	117.00	317,000,000.00	749.00	632.00	311,000,000,000.00	10,445,852.00	36.30
19		Belgium	2006	BE	135.00	446,000,000.00	954.00	819.00	327,000,000,000.00	10,511,382.00	35.70

Figure 43: Google spreadsheet