Finding The Impact of GDP on Education Attainment in Austria

A Data Management Plan created using DMPonline

Creator: Ahmad Alhirthani

Affiliation: Universitaet Wien - University of Vienna (Austria)

Template: European Commission (Horizon 2020)

ORCID iD: https://orcid.org/0000-0003-0241-0268

Project abstract:

In this project we will try to find how the education attainment in Germany is affected by the gross domestic product. This research depends on two datasets "Education Attainment for Population Aged 15-64, 1870-2010" and "GDP World Bank Data".

Last modified: 22-04-2019

Finding The Impact of GDP on Education Attainment in Austria - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

In this work we will collect data which helps in finding the impact of GDP on education attainment in Austria.

Explain the relation to the objectives of the project

The main objective of the project is to find the impact of the GDP on the educational attainment. so i collect two datasets which have a direct relation with project, since the first dataset talked about "Education Attainment" and the second one talked about GDP.

Specify the types and formats of data generated/collected

The first dataset "Education Attainment for Population Aged 15 and Over" is in .csv format. It consist of 1898 records of text and numbers.
The second one "GDP World Bank Data" is in .csv format. It contains text and numbers.
The resulted dataset is in .csv format. It contains only a numeric data.

Specify if existing data is being re-used (if any)

Question not answered.

Specify the origin of the data

The first dataset is "Education Attainment for Population Aged 15 and Over" can be found in the link: https://www.kaggle.com/ibrahimmukherjee/gdp-world-bank-data
The second one is "GDP World Bank Data" can be found in the link: http://www.barrolee.com.

State the expected size of the data (if known)

The expected data size will not consume much storage, since the processing operation is simple and it will not add any extra data to original data. So the resulted data size will not exceed the size of the collected data which is less than 2MB.

Outline the data utility: to whom will it be useful

The development of education is an essential task for politicians, so this resulted data will be helpful for politicians on make a right decisions regarding GDP having an impact on educational attainment. Thus the project will be useful to student in the Tertiary schools in austria In particular and all students in Austria generally.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Once the project is completed and the results appear dataset will be uploaded to Zenodo which is open source repository. Code and everything related to the project will be uploaded to Github.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Adding to the open-access repository, the data was reserved on Github in order to allow alternative version control using git. The reserved data link was added to the cited metadata.

Outline naming conventions used

About the resulted data, it contains 3 columns. Years intervals which refers to five years interval startting from 1965 and up to 2010, GDP amount which refers to the GDP summation for five years and Education attainment rate which refers to the rate of passing student to the enrolled student.

Outline the approach towards search keyword

In the start I searched on the internet to collect keywords that commonly used with data science and Data Management Planning then I tried to find keywords related to GDP and Education attainment in Austria.
keywords: "Data science" "DMP" "Data Management Planning" "Data Repositories" "GDP and Education", "GDP and Education in Austria", "GDP in Austria", "Education in Austria", "Tertiary Education in Austria".

Outline the approach for clear versioning

This is an educational project. That means we may do not need to handle another versions. But if we need to release versions, its easy because we are using Github which is suitable for handling versioning.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

I used Dublin Core Metadata Initiative" "DCMI" schema as standards on creating metadata.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

All resulted data will be openly available and no data will kept closed.

Specify how the data will be made available

All resulted data will be deposited in Zenodo which is helps in mading data availablity. Project files and source code will be deposited in Github which also helps in mading data availablity.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

I have collected tow datasets in the .csv format. CSV format can be viewed or edited by any text editor. So there is no required tools for accessing data.

Specify where the data and associated metadata, documentation and code are deposited

All resulted dataset will be deposited in Zenodo .Project files, metadata and source code will be deposited in Github.

Specify how access will be provided in case there are any restrictions

Zenodo and Github are trusted repostories, thus they can overcome any restrictions that anyone may face while accessing data.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

I used metadata "based on Dublin Core Metadata Initiative schema" this schema allows other users to find my project easily. Also I used csv format in generating the resulted data. Csv format can be easily interoperated by either other users or software.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

All data that I used and stored in during the project are either plain text or normal num. This makes it easy to understand and to interoperate. So it is not required to define new vocabularies.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

I will publish the resulted dataset under the licence Creative Commons Attribution 4.0 International License which means that anyone can use the published data As he pleases Without my permission

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Results will become publicly available immediately after completion working in the project and export of results.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

The resulted data will stay useable all the time because it contains historical data about GDP and education and anyone may need it in the future.

Describe data quality assurance processes

I made some computation manually to ensure that the process is working thus to preserve data quality assurance.

Specify the length of time for which the data will remain re-usable

Because results contain historical data about GDP and education, so it will be needed in the distant future.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

As mentioned before i am working in an educational project and the size of the data is less than 2MB. So i will use free hosting for my data.

Clearly identify responsibilities for data management in your project

Find techniques for data collecting which ensure data quality, accuracy and legitimacy.
Find procedures for data handling and analysis with attention to all technical aspects.
Support others in the daily use of data systems and ensure adherence to legal and company standards.
Ensure digital databases and archives are protected from security breaches and data losses.

Describe costs and potential value of long term preservation

As mentioned before i am working in an educational project and the size of the data is less than 2MB. So i will use free hosting for my data. Github save data for long time, since the project will be in the long term preservation.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

As mentioned before i am working in an educational project and we have no sensetive data but if

the size of the data is less than 2MB. So i will use free hosting for my data. Github save data for long time, since the project will be in the long term preservation.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Regarding the first datast "Education Attainment for Population Aged 15 and Over", I wrote to Professor Jong-Wha Lee by email to get his approval of the use of the datast and he accept my request. For the second dataset "GDP World Bank Data", it has a CC0 "public domain" license. Also I will refer to the source of data when publishing the resulting dataset.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Question not answered.