Financial government data analysis in austria

A Data Management Plan created using DMPonline

Creator: Oliver Steizinger

Affiliation: Other

Template: European Commission (Horizon 2020)

ORCID iD: https://orcid.org/0000-0002-9812-3820

Project abstract:

In this project, data from eurostat is used to analyze the financial movements in austria. Microsoft SQL Server is used to import and transform the data, Microsoft Power BI is used for visualisaton.

Last modified: 22-04-2019

Created using the DMPonline. Last modified 22-04-2019

1 of 4

Financial government data analysis in austria - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

We use four data sources from the european government:

- Net international investment position - quarterly data, % of GDP
  - https://data.europa.eu/euodp/en/data/dataset/Zaa87C9KaW0Oc7V1BeXRA
- Portfolio investment - quarterly data, million units of national currency
  - https://data.europa.eu/euodp/en/data/dataset/XPGtzRQI1bar3IO50yMYGg
- Direct investment - quarterly data, million units of national currency
  - https://data.europa.eu/euodp/en/data/dataset/Yu9NbJr3KNmCNuAcrrzWFQ
- General government gross debt (EDP concept), consolidated - quarterly data
  - https://data.europa.eu/euodp/en/data/dataset/uEVcriXpO2FOSnwlSkt4Q

Its quarterly data and is measured in percent of GDP of a country or in million units of national currency. It contains data from all european countries between 1995 and 2018. (not every country has data from 1995)

We use this data to compare the movements over time to find some correlation between the measurements and may help in decision processes of politicians.

Github repository:
https://github.com/Darigas1/data_stewardship_01127076

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **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?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

The data is pretty simple. There are some values (percent in GDP or million units of national currency), a time dimension, the country code and a value identifier.

We have four different data sets with similar structures to compare with.

DOI for raw data:

https://doi.org/10.5281/zenodo.2648201

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **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)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

The raw data is public data from the european union. My process and the final solution is published on github.
https://github.com/Darigas1/data_stewardship_01127076

2.3 Making data interoperable:

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

Since the raw data is a simle text file and the scripts are in SQL format, it should be interoperabe, at least inside the EU.

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **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**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Its open data and the SQL scripts can be reused.
Only Microsoft Power BI could be a problem.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Costs are low because of small data amounts. Even the analysis would be much better with more data. For now there are to little datapoints to make some forecasts.
The EU is resposible to give access to the data.

4. Data security

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

Because the data used is open data, there are not much concerns about data security.
The hole project is MIT Licensed and can be freely used.

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

Its possible to make assumptions without the hole "picture". That means there could be bad decisions for people because of the interpretation of the data! A professional opinion is needed for further research.

6. Other

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

None

Created using the DMPonline. Last modified 22-04-2019

4 of 4