Financial government data analysis in austria

A Data Management Plan created using DMPonline

Creator: Oliver Steizinger

Affiliation: Other

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

Template: European Commission (Horizon 2020)

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

Financial government data analysis in austria - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

Analyze financial government data in austria

Explain the relation to the objectives of the project

Its all about finance in europe

Specify the types and formats of data generated/collected

Simple text files are downloaded and imported

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

Data from the EU is used:

- 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

Specify the origin of the data

Financial reports from each country in the EU

State the expected size of the data (if known)

Very low (a few hundred KB)

Outline the data utility: to whom will it be useful

It is usefull to politicians or people interested in european finance.

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

Outline the discoverability of data (metadata provision)

There is not much metadata used.
It could be found on https://data.europa.eu

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?

DOI is used for data and source code:
Data - https://doi.org/10.5281/zenodo.2648201

Code - https://doi.org/10.5281/zenodo.2648674

Outline naming conventions used

Not needed

Outline the approach towards search keyword

Not needed, not complex enought

Outline the approach for clear versioning

Versioning on github

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

Not that much metadata in this project.

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

Everything is avaiable on github.
https://doi.org/10.5281/zenodo.2648674

Specify how the data will be made available

On github with DOI

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)?

To reproduce the hole project, Microsoft Power BI is needed. The desktop version can be freely downloaded onhttps://powerbi.microsoft.com/de-de/
Also some SQL-Server will be necessary

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

All on GitHub

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

As long as the EU exists, this should be accessible.

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.

The provided Power BI file on Github has the data included and can be used immediately for research and can be improved by other data scientists. Should be easy to use.

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?

Microsoft Power BI provides some interfaces for interoperability. For example R can be used in Power BI to create some forecasts or other statistical methods.

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

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

Its open data, everybody can use it.

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

After the EU updates their data

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

Open data

Describe data quality assurance processes

Check out by my own. In real life this should do some expert.

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

As long a github exists

3. Allocation of resources

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

Its a lot of work and nobody covers it.

Clearly identify responsibilities for data management in your project

Oliver Steizinger is responsible for everything

Describe costs and potential value of long term preservation

Not that much because quarterly data preservation doesnt need a lot of space and maintainance.

4. Data security

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

No sensitive data 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)

No