RELATION BETWEEN NUMBER OF UNEMPLOYMENT PEOPLE IN NEW YORK AND NUMBER OF DNA STRUCTURES RELEASED BY PROTEIN DATA BANK

A Data Management Plan created using DMPonline

Creator: IBAIDA TAHA, AYSAR ISHTAWI

Affiliation: Other

Template: Digital Curation Centre

ORCID iD: https://orcid.org/0000-0001-6592-7016

Project abstract:
Our project presents the relation between Number of Unemployment people in New York and Number of DNA Structures Released by protein data bank from 2000 to 2018, It contains a data sets from two different open data repository.

Last modified: 11-06-2018

RELATION BETWEEN NUMBER OF UNEMPLOYMENT PEOPLE IN NEW YORK AND NUMBER OF DNA STRUCTURES RELEASED BY PROTEIN DATA BANK

DATA COLLECTION

What data will you collect or create?

Data that we used in our experiment was experimental observation and samples from the research done by protein data bank and US government's open data repository, This is the first file https://www.rcsb.org/stats/growth/dna, doi: 10.1038/nsb1203-980, and has volume (1 KB) and the second is https://catalog.data.gov/dataset/local-area-unemployment-statistics-beginning-1976, Date and time accessed on 24/3/2018 at 17:30, and has a volume (3,5 KB), both of them and the result file also will be saved as comma separated value files (.csv).

How will the data be collected or created?

We took datasets from these two different open data repository:

- Unemployment data from the U.S. Government's repository (https://www.data.gov)
- DNA structures from Protein Data Bank (https://www.rcsb.org)

After that, we converted these files to excel files and we represented the relation between them in a chart using excel software.

DOCUMENTATION AND METADATA

What documentation and metadata will accompany the data?

First Data source is Data.gov and it's follows the Project Open Data schema – a set of required fields (Title, Description, Tags, Last Update, Publisher, Contact Name, etc.) for every data set displayed on Data.gov.
Second Data source is Protein data bank and the metadata schema used is the one used by the repository.

ETHICS AND LEGAL COMPLIANCE

How will you manage any ethical issues?

No ethical issues (no human subject data collected).

How will you manage copyright and Intellectual Property Rights (IPR) issues?

First, about copyright, the data we used from protein data bank and U.S. Government's repository is public and accessible for everyone.
Second, the data is owned by data bank and U.S. Government's repository and it's used for further non-commercial or commercial dissemination.

STORAGE AND BACKUP

How will the data be stored and backed up during the research?

Digital copies of the result will be stored on Figshare (https://figshare.com)
The stability and accessibility of Figshare provides a suitable option for long-term storage of data and should ensure minimal risk of data loss.
How will you manage access and security?

All data and metadata will be stored privately in the cloud on Figshare until publication, after which point it will be made open-access. Data will be searchable on Figshare, and downloadable by any user.

SELECTION AND PRESERVATION
Which data are of long-term value and should be retained, shared, and/or preserved?

The final dataset will be transferred to the Figshare repository, which ensures sustainable archiving of the final research data.
What is the long-term preservation plan for the dataset?

For long-term archiving, we used Figshare repository, and it didn't cost us any fees, since it's a free open access digital repository, the uploading process was simple and didn't need a lot of effort and time.

DATA SHARING
How will you share the data?

Potential data users for the data that we will compile include people who interested in statistical and statistics.
We used Figshare to publish the data and this is the output.
Doi: https://doi.org/10.6084/m9.figshare.6480854.v1
Are any restrictions on data sharing required?

No restrictions on data necessary, or ethical or privacy issues. Data will be free to use under the expectation that it will be correctly attributed and cited using the Figshare DOI.

RESPONSIBILITIES AND RESOURCES
Who will be responsible for data management?

Team members will be responsible for data management.
What resources will you require to deliver your plan?

No resources needed.