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

Affiliation: Other

Template: Digital Curation Centre

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

Project abstract:

This report 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: 24-05-2018

Copyright information:

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, The first file has volume (1kB) and the second is (3,477 kB), The excel spreadsheet will save as a comma separated value (.csv) file.

How will the data be collected or created?

We took a data sets from this 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)

Documentation and Metadata

What documentation and metadata will accompany the data?

All data has been worked in comma separated value (.csv) format with accompanying meta data file in text (.txt) format.

Documentation, tutorials, and metadata for using both data and software will be archived with the relevant product. Examples of data repositories we will use include Figshare. Examples of software repositories include Zenodo and Figshare. The goal of these combined efforts is to maintain archives of all data and software products indefinitely.

Ethics and Legal Compliance

How will you manage any ethical issues?

No ethical issues (no human subject data collected), will be protected by patent before or after publication. At the time of publication supporting data will be made available freely, or under requested license, according to sensitivity.

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?

USB hard drive and paper copies of data and metadata to be stored. Digital copies will be stored on Figshare (https://figshare.com).
The 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 under a Creative Commons license, and citable in its own right. 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?

All data are publicly available and preserved.

What is the long-term preservation plan for the dataset?

For long-term archiving, there are a couple of options in place. like we will save the data by upload it to open data repository like Fedora, Dspace, or Ckan to keep it.

In addition, For short-term archiving purposes this data is backed up nightly and retained for 30 days, and each of the team members has a copy of the data, and we upload data to Google Drive.

Data Sharing

How will you share the data?

Potential data users for the data we will compile include people who interested in statistical and statistics and include anyone working with long-term data.

All data will be published using the CC0 Public Domain Dedication if allowed by the repository, with CC-BY as a fallback if the repository does not allow CC0.

Are any restrictions on data sharing required?

No restrictions on data necessary, or ethical or privacy issues. Intellectual property rights will rest with the original author of the data (Ibaida), and the project supervisors (Aysar, Ibaida). 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?

The team members will be primarily responsible for managing all information as relates to this project according to the FAIR principles.

What resources will you require to deliver your plan?

No resources needed.