Version 2

Jun 06, 2021

Data collection V.2

Alise J Ponsero[1]

[1]University of Arizona

1 Works for me Share This protocol is published without a DOI.

CURE_BAT102
Tech. support email: aponsero@email.arizona.edu

Alise Ponsero
University of Arizona

ABSTRACT

This protocols aims to details the steps necessary to collect new data points for the "Database Justice League" project.

PROTOCOL CITATION

Alise J Ponsero 2021. Data collection. **protocols.io**
https://protocols.io/view/data-collection-bvjpn4mn
Version created by Alise Ponsero

CREATED

Jun 06, 2021

LAST MODIFIED

Jun 06, 2021

PROTOCOL INTEGER ID

50511

MATERIALS TEXT

DB_DataCollection_TD.csv

DISCLAIMER:

This protocol is for teaching purpose

BEFORE STARTING

Before starting this protocol, you need to have recieved the list of articles you will be working on.
Contact the professor if you haven't received the dataset.

Download data

1 Once you have received the link to download your dataset, download the csv file locally.
Make sure to save it on your computer in a convenient folder!

In this protocol we will perform the data collection on a test dataset:

DB_DataCollection_TD.csv

2 As you can see in this test dataset, some preliminary information have been collected automatically such as the article name, authors and publication date.

However some information will need additional efforts to be collected, such as the database name, link and availability.

Database name and link

3 In the dataset, you'll find a column name "DOI", this is a usique identifier given to all scientific publications. This allows to easily find the article online.

In order to access the paper search in you web browser :
https://doi.org/ + DOI number

In the first example it would be: https://doi.org/10.1093/nar/gkaa818

4 The first information you need to collect is the name of the database! It seem trivial, but people use acronyms and are often not consistent in naming databases. So be careful!

There is mainly two possibilities: either the article is presenting a new database or it is an article about an update of an already published database

Step 4 includes a Step case.
New database
New version of a database

Database URL

step case

New database

If your article is about a new database, you'll need to look at the title and abstract to find the name of the database.

In the first example of the test set, the abstract of the article statrs by "We introduce the Nucleome Data Bank (NDB), a web-based platform to simulate and analyze the three-dimensional (3D) organization of genomes."
Since the database name has two version, an acronym and a full name, you'll need to report both in the column "**resource_name**"

Here it would be:
NDB: Nucleome Data Bank

If only one name exists, simply report the name in the "ressource_name" column

5 Next, we need to find the database URL. In general, the URL is mentionned in the article abstract.

As an example for the first article in the test dataset, you see in the article abstract:
"The NDB aims to be a shared resource to biologists, biophysicists and all genome scientists. The NDB is available at https://ndb.rice.edu."

Here you need to report the full URL (https://ndb.rice.edu) under the "current_access" column.

Database accessibility

6 Finally, we need to assess if the URL for the database is working. Click on the URL you identified!

Two possibilities: The URL is working or not!
Step 6 includes a Step case.
URL working
URL not working