



Version 1

Jun 06, 2021

# Data collection V.1

Alise J Ponsero<sup>1</sup><sup>1</sup>University of Arizona

1

Works for me



Share

This protocol is published without a DOI.

CURE\_BAT102

Tech. support email: [aponsero@email.arizona.edu](mailto: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)  
<https://protocols.io/view/data-collection-bvjmn4k6>



## LICENSE

This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

## CREATED

Jun 06, 2021

## LAST MODIFIED

Jun 06, 2021

## PROTOCOL INTEGER ID

50509

## 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 unique 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**