MWP5 Final Project: 5 findability opportunities to access the portfolio dataset

Anastasiia Iarkaeva[1]

[1]Humboldt Universität Berlin
Anastasiia Iarkaeva: Student

Apr 15, 2022

1

dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

Anastasiia Iarkaeva

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

This protocol was created as part of the Information Science Master's program project MWP5 Digital Information Management at the Humboldt University of Berlin.

It aims to help a potential user group of the decision-making stakeholders to find the examined dataset for cluster modeling (https://doi.org/10.6084/m9.figshare.12743639.v1).

Overall five opportunities are both described textually and shown visually.

In all cases, except the DOI system, the search results can be reduced by applying multiple filters to find the wanted dataset faster and more precisely.

DOI

dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

Citation: Anastasiia Iarkaeva MWP5 Final Project: 5 findability opportunities to access the portfolio dataset https://dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

Anastasiia Iarkaeva 2022. MWP5 Final Project: 5 findability opportunities to access the portfolio dataset . **protocols.io**
https://dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

dataset findability, pubmed literature model, data structures, figshare

protocol ,

Mar 30, 2022

Apr 15, 2022

60094

:

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

DOI

1 Use the **DOI** to directly access the dataset.

A DOI is a Digital Object Identifier. That is a Digital Identifier of an Object, not an Identifier of a Digital Object.

DOIs allow objects to be identified and accessed with certainty. Once DOIs are allocated, they never change for that object (Driven by DOI, n.d.).

Driven by DOI. (n.d.). Retrieved March 30, 2022, from https://www.doi.org/driven_by_DOI.html

Citation: Anastasiia Iarkaeva MWP5 Final Project: 5 findability opportunities to access the portfolio dataset https://dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

1.1 For that, open the **DOI system** on doi.org, add "/" sign and the unique identifier of the dataset (doi: 10.6084/m9.figshare.12743639.v1) or the data descriptor article (doi: 10.1038/s41597-020-00749-y) as shown in the video.

You will be navigated to the requested source at once.

1.2 Another opportunity to find the dataset in the **DOI system** is to use the search field by typing in or pasting the identifier as shown in the video.

Figshare

2 Use figshare repository where the data is located to find the dataset.

For that, open the repository website and type or paste the keywords *pubmed* and *literature* which were selected from the dataset title *STS Model of the PubMed Literature*.

The results can be filtered and sorted to reduce the amount of found datasets and to find a proper match, e.g. from 7109 to 2437 datasets here.

Tip: use a "+" sign to reduce the number of found results as it serves the purpose of exact matching.

Dimensions

3 Use Dimensions Abstracting & Indexing database where the datasets from figshare are harvested automatically.

- From the home page access right away the web retrieval system.
- Use the keywords *pubmed literature* in the search bar.

Citation: Anastasiia Iarkaeva MWP5 Final Project: 5 findability opportunities to access the portfolio dataset https://dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

- Then the results can be filtered by multiple parameters, e.g. type of data, repository name, year, etc. Here the repository name filter was applied - the number was reduced from 2732 to 30 results.
- The third hit is the wanted dataset.

If searching the keywords "*pubmed literature model*" the hit will appear already on the first page (6th here).

Google dataset search

4 Use Google dataset search which collects a lot of well-curated datasets' metadata if they conform to the standards defined by the schema.org consortium.

- For that, open the Google dataset search engine and type or paste the keywords *pubmed* and *literature*.
- When you type, it will show a dropdown with the best match results that you can open right away as it is shown in the video below.
- If you press the "enter" key it will show a page with search results that can be filtered to find a proper dataset.

PubMed Central

5 Searching via PubMed Central with the keywords "*pubmed literature model*" you will hit the data descriptor article as the first result.

Citation: Anastasiia Iarkaeva MWP5 Final Project: 5 findability opportunities to access the portfolio dataset https://dx.doi.org/10.17504/protocols.io.x54v9yr2qg3e/v1

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,