



APR 12, 2024

## IRIS Software Protocol

Leonardo Zilli<sup>1</sup>, Erica Andreose<sup>1</sup>, Salvatore Di Marzo<sup>1</sup>

<sup>1</sup>University of Bologna



Leonardo Zilli  
University of Bologna

### ABSTRACT

This project aims to investigate the discoverability of University of Bologna's scholarly output within OpenCitations Meta, a platform that stores and delivers bibliographic metadata for all the publications involved in OpenCitations Index which contains recording citations allowing the user to visualize all of the citation links between a document and another. Specifically, this project aims to analyze the coverage of University of Bologna's publications within the OpenCitations Meta, particularly those deposited in the IRIS institutional repository. Various publication types are observed to discern patterns in representation. Utilizing OpenCitations data, the citation impact of these IRIS publications is quantified, encompassing both the citations they receive and those they provide. Additionally, the study delves into the nature of citations, distinguishing between those involving publications within IRIS and those referencing external sources. The study unveils the citation impact of IRIS publications, clarifying their influence within the scholarly community. By dissecting citations, the research also delineates the interplay between internal and external citations, providing a deeper understanding of the institution's scholarly ecosystem. The methodology used will be all compliant to the core values of Open Science, the data and the software used will all be made available to allow the user to replicate, reproduce and validate the conclusions drawn by the end results of the project, that will hold significance for researchers and academic institutions, facilitating informed decision-making and fostering a deeper understanding of scholarly communication dynamics.

### GUIDELINES

To let the user with complete reproducibility of the protocol, links to the data used are provided here.

Download link for the UNIBO IRIS bibliographic data dump, dated 14 March 2024:

<https://amsacta.unibo.it/id/eprint/7608/>

OpenCitations Meta can be accessed and queried through a REST API:

<https://w3id.org/oc/meta/api/v1>

### OPEN ACCESS



#### DOI:

[dx.doi.org/10.17504/protocols.io.3byl497wjgo5/v1](https://dx.doi.org/10.17504/protocols.io.3byl497wjgo5/v1)

**Protocol Citation:** Leonardo Zilli, Erica Andreose, Salvatore Di Marzo 2024. IRIS Software Protocol. [protocols.io](https://dx.doi.org/10.17504/protocols.io.3byl497wjgo5/v1) <https://dx.doi.org/10.17504/protocols.io.3byl497wjgo5/v1>

**License:** 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, provided the original author and source are credited

**Protocol status:** In development  
We are still developing and optimizing this protocol

**Created:** Apr 12, 2024

Last Modified: Apr 12, 2024

PROTOCOL integer ID: 98117

Funders Acknowledgement:

University of Bologna  
Grant ID:  
<https://ror.org/01111rn36>

BEFORE START INSTRUCTIONS

Before starting, we suggest to make sure you have Python3.x installed on your computer, in addition, in order to correctly execute the provided scripts, you must install the required libraries:  
[requirements.txt](#)  
- pandas

IRIS Dataset Preparation

1

Dataset

UNIBO IRIS bibliographic data dump, dated 14 March 2024<sup>NAME</sup>

<https://amsacta.unibo.it/7608/1/iris-data-2024-03-14.zip> <sup>LINK</sup>

After downloading the IRIS dataset, the contents appear divided into 7 different csv files, descriptive as the following:

- "ODS\_L1\_IR\_ITEM\_CON\_PERSON.csv": information about the people involved in the publications (authors, editors, etc.)
- "ODS\_L1\_IR\_ITEM\_DESCRIPTION.csv": the string of the authors and other related metadata of publications
- "ODS\_L1\_IR\_ITEM\_IDENTIFIER.csv": the identifiers (including DOIs) of publications
- "ODS\_L1\_IR\_ITEM\_LANGUAGE.csv": the language in which the publication has been written (when applicable)
- "ODS\_L1\_IR\_ITEM\_MASTER\_ALL.csv": basic metadata information of publications (title and date of publication)
- "ODS\_L1\_IR\_ITEM\_PUBLISHER.csv": the publishers of publications
- "ODS\_L1\_IR\_ITEM\_RELATION.csv": additional metadata related to the context of publications (publication venue, editors, etc.)

These files are connected among themselves with a unique item ID tied to each entry, after we gathered the files we converted them into dataframes with Python Pandas.

## Software

<b>python</b>	NAME
3.6	OS
Guido van Rossum	DEVELOPER

(<https://pandas.pydata.org/>)

## 2 The csv files are converted into Dataframes

```
iris_id_df = pd.read_csv('ODS_L1_IR_ITEM_IDENTIFIER.csv')
iris_master_df = pd.read_csv('ODS_L1_IR_ITEM_MASTER_ALL.csv')
iris_relation_df = pd.read_csv('ODS_L1_IR_ITEM_RELATION.csv')
```

## 3 The DataFrames are grouped into one mashup dataframe through the item\_ID, to see the various connections and data that the files could offer for the same entry.

```
iris_grouped_df = pd.merge(iris_id_df, iris_master_df, iris_relation_df,
on='ITEM_ID')
```

## OpenCitations Meta Dataset Querying

## 4

### Dataset

<b>OpenCitations Meta SPARQL endpoint</b>	NAME
<a href="https://opencitations.net/meta/sparql">https://opencitations.net/meta/sparql</a>	LINK

The OpenCitations Meta Sparql endpoint allowed us after querying it, to download a csv file of the results of our queries which we then transformed into dataframes.

These are the queries that gave as a result the files:

```
# 1 Sparql query
```

```
# 2 Sparql query
```

```
# 3 Sparql query
```

**4.1** We then use these csv files to create the pandas DataFrames that will be used to perform the analyses.

```
meta_df = pd.read_csv(f'meta.csv')
```

## OpenCitations Index Dataset Querying

5

Dataset

OpenCitations Index

NAME

<https://opencitations.net/index/sparql>

LINK

The OpenCitations Index Sparql endpoint allowed us after querying it, to download a csv file of the results of our queries which we then transformed into dataframes.

These are the queries that gave as a result the files:

```
# 1 Sparql query
```

```
# 2 Sparql query
```

```
# 3 Sparql query
```

**5.1** We then use these csv files to create the pandas DataFrames that will be used to perform the analyses.

```
index_df = pd.read_csv(f'index.csv')
```

## Research Question answering

**6** We answer the questions:

**6.1** ▪ *What is the coverage of the publications available in IRIS (strictly concerning research conducted within the University of Bologna) in OpenCitations Meta?*

```
filter the IRIS dataframe for publications conducted at UNIBO
check the intersection of the DOIs from the iris_df with the
meta_df
```

**6.2** ▪ *Which are the types of publications that are better covered in OpenCitations Meta?*

checking the type of the resulting data of the first question

- 6.3
- *What is the amount of citations (according to OpenCitations Index) coming from the IRIS publications that are involved in OpenCitations Meta (as citing entity and as cited entity)?*

checking the number of entities coming from IRIS, stored in OpenCitations Meta, that cite or are cited inside of OpenCitations Index

- 6.4
- *How many of these citations come from and go to publications that are not included in IRIS?*

check the results of the third question to find which citations cite entries not available in IRIS, and which entries in IRIS cite entries not available in IRIS

- 6.5
- *How many of these citations involve publications in IRIS as both citing and cited entities?*

check the results of the third question to find how many citations in OpenCitations Index involve the publications in IRIS that are both acting as citing and cited

## Data Visualization

- 7
- The visualizations of the results obtained are computed with the `__` library.

#### Note

*What is the coverage of the publications available in IRIS (strictly concerning research conducted within the University of Bologna) in OpenCitations Meta?*

*-Insert graph*

#### Note

*Which are the types of publications that are better covered in OpenCitations Meta?*

*-Insert graph*

#### Note

*What is the amount of citations (according to OpenCitations Index) coming from the IRIS publications that are involved in OpenCitations Meta (as citing entity and as cited entity)?*

*-Insert graph*

#### Note

*How many of these citations come from and go to publications that are not included in IRIS?*

*-Insert graph*

#### Note

*How many of these citations involve publications in IRIS as both citing and cited entities?*

*-Insert graph*

