




IMFDataPy: A Python package for IMF Data Discovery and Extraction

Irina Klein^{1*} and Sou-Cheng Choi^{1,2*}

¹ Illinois Institute of Technology, USA ² SAS Institute Inc., USA * These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

Summary

The IMF's RESTful APIs for retrieving international economic data series are technically involved. The structure of the returned JSON datasets varies from one series to another. It is too complex for an unprepared or non-technical user to download or use the data.

We have created an installable Python package, **IMFDataPy**, for discovering and extracting IMF data series. It serves to mask or wrap the complexities of the IMF JSON RESTful API so that the users would not need to handle JSON data or its underlying metadata. It provides an intuitive way for users to search through the series or indicator names.

IMFDataPy is designed with an extensible software architecture and simple APIs. It contains comprehensive and searchable documentation; as well as unit tests that ensure the functionality of the package as designed. The package also contains robust input parsing; provides summaries of the downloaded IMF time series data and metadata; and last but not least, gives a number of economic and financial use cases of **IMFDataPy**.

The current work is released as a Python package on Python Package Index, [IMFDataPy PyPI page](#) and the source code is available at the GitHub repository, [IMFDataPy GitHub page](#).

Statement of need

Overview

The International Monetary Fund (IMF) is an international organization that provides financial assistance and advice to its 190 member countries out of 195 countries in the world. Apart from advising services, the IMF periodically collects large amounts of data on various economic indices from its member countries.

An IMF data series (e.g., International Financial Statistics) is a dataset containing multiple economic indicators. An economic indicator (e.g., Gross Domestic Product) is a set of time-indexed numeric values that represents an economic index or metric. **Dimension** refers to the metadata that pertains to all indicators within a specific IMF series. Here, metadata is defined as a collection of information that provides descriptive and structural details about the data itself. The dimension most commonly include the following items:

- Area (e.g., the US) — or in some datasets called Country — may refer to a territorial entity of a state understood by international law, a territorial group (e.g., Eurozone), or a non-sovereign territory for which statistics are maintained;
- Frequency (e.g., quarterly);
- Period (e.g., from 2020 to 2022).

37 Table 1 lists some of the more than 30 monthly, quarterly, or annual data series for 190 member
38 countries from the 1960's to present available from IMF ([IMF, 2023a](#)).

Table 1: Examples of IMF data series and indicators.

Series	Series Name	Examples of Indicators
IFS	International Financial Statistics	Gross Domestic Product, Interest Rates, Unemployment Rates, Consumer Price Index, Industrial Production, Exchange Rates, Export and Import, Government Revenues and Expenditures
GFS	Government Finance Statistics	Financial assets and liabilities classified by sector, Government Revenue, Government Cash Flow
HPDD	Historical Public Debt Database	Debt To GDP Ratio
PCPS	Primary Commodity Price System	Indices of market price for fuel and non-fuel commodities
DOTS	Direction of Trade Statistics	Value of Imports and Exports, Value of Trade Balance
FDID	Financial Development Index Database	Financial Development Index, Financial Market Index
CPI	Consumer Price Index	CPI for various goods and services groups

39 There are four main data extraction methods, which we summarize below. For their detailed
40 IMF help and documentation, see ([IMF, 2020a](#)).

- 41 ■ **Web Query Interface** allows user interactions and customization of data tables and
42 graphs online.
- 43 ■ **Bulk Download** allows downloading a zip file containing csv files for data and metadata
44 for each dataset ([IMF, 2023c](#)).
- 45 ■ **IMF Data Mapper** and **IMF Mobile App** provides data lists, summaries, and visualization
46 for some of the IMF indices. Data such as real GDP trends can be accessed through a
47 web browser or the IMF mobile app on a phone or a tablet ([IMF, 2015, 2023b](#)).
- 48 ■ **JSON RESTful Web Service API** can be programmatically accessed using Python, R,
49 or other programming languages to download JSON files automatically. Briefly, **JSON** is
50 a file format for data storage and transmission that consists of attribute-value pairs and
51 arrays. It is readable by machines, but not readily comprehensible by people. **RESTful**
52 **API** is an interface that enables clients (programs or devices) to interact with server
53 resources using the **REpresentational State Transfer** design pattern. This design pattern
54 is based on the exchange of client requests and server responses as described in Chapter
55 5 of ([Fielding, 2000](#)).

56 While powerful and flexible, the RESTful API is however not exactly intuitive for first-time or
57 less technical users. The main issue is that the data is stored in datasets called series, with
58 each series associated with multiple dimensions. For example, to find a country's Consumer
59 Price Index (CPI) or Gross Domestic Product (GDP), one would need to first discover the
60 correct names of the containing series and dimensions, followed by a text search of well-selected
61 keywords. ([Dew, 2016, 2018](#)) illustrate the complexity well. Additional examples are available
62 in our [IMFDataPy Github repository](#) (in the folder 'demo').

@FREQ	Q
@REF_AREA	US
@INDICATOR	NGDP_R_SA_XDC
@UNIT_MULT	6
@BASE_YEAR	2012
@TIME_FORMAT	P3M

Obs[0]	Obs[1]	Obs[2]			
@TIME_PERIOD	2022-Q1	@TIME_PERIOD	2022-Q2	@TIME_PERIOD	2022-Q3
@OBS_VALUE	4981022	@OBS_VALUE	4973817.8	@OBS_VALUE	5005430.3

Figure 1: Data Structure of Quarterly Real Gross Domestic Product observations for the US in 2022.

Example 1. *Real Gross Domestic Product (RGDP)* is the inflation-adjusted monetary value of goods and services in a country during a specific time period. *Quarter-on-quarter (QoQ)* is a change in value between one quarter and the previous quarter. To access QoQ RGDP data using the JSON RESTful API, we need to follow two steps:

- First, understand or look up the information of its containing series by using the `DataSet` method, and specify the necessary dimensions.
- Second, request the data using the `CompactData` method.

Description of the above and other methods can be found in our [GitHub demo](#) and the IMF website (IMF, 2020b). We refer readers to our [Python Jupyter Notebook Example](#) for step-by-step implementation details.

Existing Software Packages

IMFDataPy is our Python package that allows users to retrieve data directly from the IMF's servers. In Table 2, we compare **IMFDataPy** with other R or Python software packages that have similar functionality.

Table 2: Existing packages to explore and extract the IMF data from IMF's servers.

Name (Language)	Functionality	Available Series	Limitations
IMFData (Lee, 2019) (R)	Search through series and index codes and output data, given series name, index codes and other parameters	All	Not actively maintained. Removed from CRAN. An archive version can be installed.
imfr (Gandrud, 2020) (R)	Load data given series name, index code and other parameters	All	Not actively maintained. Removed from CRAN. An archive version can be installed.
datapungi (Otterson, 2020) (Python)	Load data & metadata given series name, index code and other parameters	All	Some series (other than IFS) resulted in errors. Little documentation was provided. No information on unit tests. Source code was not available.

Name (Language)	Functionality	Available Series	Limitations
imfpy (Kearney, 2021) (Python)	Search through the datasets, download data into a Pandas DataFrame, visualize data	DOTS	Only one IMF series.
PyIMF (Eggers, 2022) (Python)	Search through datasets and indices and output data, given index codes and other parameters	All	Installation using pip results in error. No source repository found. Documentation was not provided. No unit tests.
IMFDataPy (IMFDataPy?) (Python)	Search through series names. Download data & metadata given series name and index search terms into Pandas DataFrames and CSV files	AFRREO, BOP, DOT, FSI, GFSR, HPDD, IFS	Limited number of series (more to be added). Limited built-in visualization of the results.

DBnomics (DBnomics team, 2023) is another noteworthy service that collects and stores economic data from the IMF (and some other 88 providers at the time of writing) on its own servers that is not associated with IMF. It also provides a user-friendly web interface for analyzing essential indices with detailed information on the data sources. Moreover, there are R and Python clients available that enable users to retrieve the data automatically.

IMFDataPy Package

IMFDataPy is designed with an object-oriented architecture, in which child objects may inherit or override the behavior of the parent object. Each of the IMF's datasets is a child of the parent class, IMF object, which in turn is a child of the abstract class, Series (Figure 2).

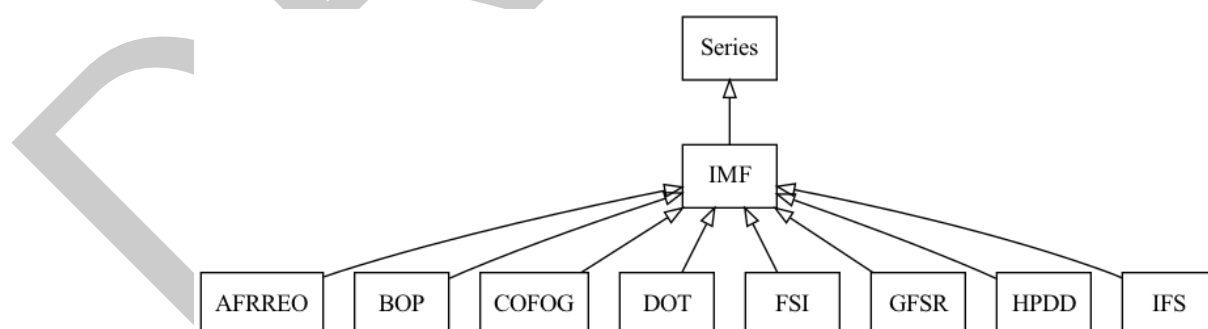


Figure 2: Object-oriented design of IMFDataPy.

All data is downloaded into a Pandas DataFrame, which is a two-dimensional tabular data format with columns of various data types that come with the Python package, Pandas. In order to produce the final DataFrame, the package first sends a request to the IMF's server to get the dimensions for a specific series and then, having analyzed which of the indicators match the search terms given by the user, the IMFDataPy package sends a request to load the data given the dimensions, including the names of the indicators (Figure 3).

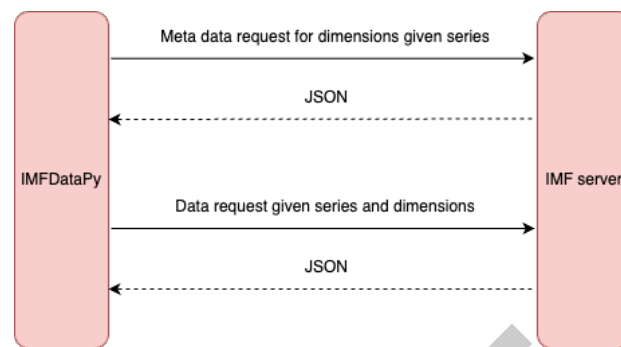


Figure 3: Sequence of requests from **IMFDataPy** to IMF's server

92 Having received the JSON file with the data, the package then transforms it into a more
93 readable Pandas DataFrame and provides additional summary statistics and visualization.

94 To install the package, a user may simply run the command:

95 `pip install imfdatapy`

96 **Example 2.** For the example of quarterly RGDP change using the **IMFDataPy** package, the
97 essential Python code is shown below. For details, we refer readers to our [Python Jupyter](#)
98 [Notebook Example](#) for details.

```

99 from imfdatapy.imf import *
100
101 ifs = IFS(search_terms=["gross domestic product"], countries=["US"],
102           period='Q', start_date="2010", end_date="2023")
103
104 df = ifs.download_data()
  
```

105 We trust with **IMFDataPy**, more people, especially beginning or non-technical users, are able
106 to access IMF data more readily.

107 Acknowledgements

108 We thank Joshua Herman and Aleksei Sorokin for discussion. We also thank Karan Bakshi
109 and Sahil Malhotra for their contribution to the repository. The authors are grateful for the
110 opportunities of presenting **IMFDataPy** in the conferences [PyData Global 2022](#); [Midwest](#)
111 [Machine Learning Symposium 2023](#); and [The 15th R/Finance Conference for Applied Finance](#)
112 [Using R](#).

113 References

- 114 DBnomics team. (2023). *DBnomics*. <https://db.nomics.world>.
- 115 Dew, B. (2016). *Machine reading IMF data: Data retrieval with Python*.
- 116 Dew, B. (2018). *IMF API with Python: An example*. <https://www.bd-econ.com/imfapi1.html>.
- 117 Eggers, C. (2022). *PyIMF*. <https://pypi.org/project/PyIMF/>.
- 118 Fielding, R. T. (2000). *Architectural styles and the design of network-based software architec-*
119 *tures*. University of California, Irvine.
- 120 Gandrud, C. (2020). *imfr*. <https://github.com/cran/imfr>.
- 121 IMF. (2015). *IMF VIDEOS. IMF new mobile app available now*. <https://www.imf.org/en/Videos/view?vid=3984275247001>.
- 122

- 123 IMF. (2020a). *IMF help documents and references*. [https://datahelp.imf.org/knowledgebase/](https://datahelp.imf.org/knowledgebase/topics/69748-help-documents-and-references)
124 [topics/69748-help-documents-and-references](https://datahelp.imf.org/knowledgebase/topics/69748-help-documents-and-references).
- 125 IMF. (2020b). *JSON RESTful web service*.
126 <https://datahelp.imf.org/knowledgebase/articles/667681-json-restful-web-service>.
- 127 IMF. (2023a). *Available datasets listed alphabetically*.
128 [https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sld=](https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sld=1479329132316)
129 [1479329132316](https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sld=1479329132316).
- 130 IMF. (2023b). *IMF Data Mapper*. <https://www.imf.org/external/datamapper/>.
- 131 IMF. (2023c). *International Financial Statistics (IFS)*. [https://data.imf.org/?sk=](https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b)
132 [4c514d48-b6ba-49ed-8ab9-52b0c1a0179b](https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b).
- 133 Kearney, L. T. (2021). *imfpy*. <https://pypi.org/project/imfpy/>.
- 134 Lee, M. (2019). *IMFData*. <https://github.com/mingjerli/IMFData>.
- 135 Otterson, J. (2020). *datapungi_imf*. https://pypi.org/project/datapungi_imf/.

DRAFT