IMFDataPy: A Python Package for IMF Data Discovery and Extraction

Irina Klein [1] Sou-Cheng Choi[1, 2]

[1]Illinois Institute of Technology, USA [2]SAS Institute Inc., USA

Data Available at IMF

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.

Series	Dataset Name	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
DOTS	Direction of Trade Statistics	Value of Imports and Exports, Value of Trade Balance

Table 1. Partial list of monthly, quarterly, or annual data series for 190 member countries from the 1960's to present available from IMF [**?**].

Data Structure

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 includes the following items:

- **Area** (e.g., the US);
- **Frequency** (e.g., quarterly);
- **Period** (e.g., from 2020 to 2022).

How to Access the Data?

1. **Web Query Interface** allows user interactions and customization of data tables and graphs online.
2. **Bulk Download** allows downloading a zip file containing CSV files for data and metadata for each dataset.
3. **IMF Data Mapper** and **IMF Mobile App** provide data lists, summaries, and visualization for some of the IMF indices.
4. **JSON RESTful Web Service API** can be accessed using Python or R to download JSON files automatically.

JSON RESTful Web Service API

JSON is a file format for data storage and transmission that consists of attribute-value pairs and arrays. It is readable by machines, but not readily comprehensible by people. **RESTful API** is an interface that enables clients (programs or devices) to interact with server resources using the **RE**presentational **S**tate **T**ransfer design pattern.

JSON Example

Example 1. A JSON string: first value in the area 'dimension' from the IFS data series.

{'@value': 'AF', 'Description':{'@xml:lang': 'en', '#text': 'Afghanistan'}}

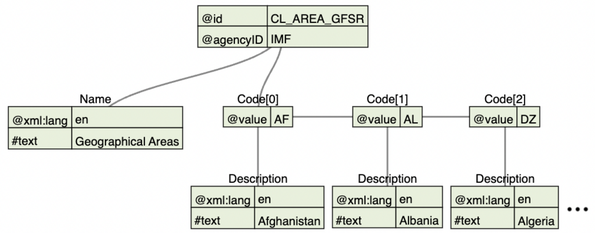

Figure 1. GFSR* Area Dimension: first three items. Note: This diagram is produced with JSON Tree Visualizer [**?**].

* Government Financial Statistics: Revenue

Existing Challenges

While powerful and flexible, the RESTful API is however not exactly intuitive for first-time or less technical users.

To access 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 **DataStructure** method, and specify the necessary dimensions.
- Second, request the data using the **CompactData** method.

A description of the above and other methods can be found in our GitHub demo and the IMF website [**?**].

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.

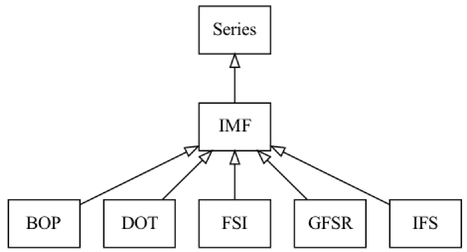

Figure 2. Object-oriented design of **IMFDataPy**.

Data Loading

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.

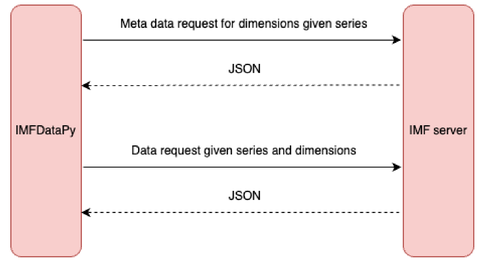

Figure 3. Sequence of requests from **IMFDataPy** to IMF's server.

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

Installation and Usage

To install the package, a user may simply run the command: `pip install imfdatapy`

Example 2. To retrieve quarterly RGDP data using the IMFDataPy package, the essential Python code is shown below.

```
from imfdatapy.imf import *

ifs = IFS(search_terms=["gross domestic product"], countries=["US"],
period='Q', start_date="2010", end_date="2023")

df = ifs.download_data()
```

References

Acknowledgements

We thank Joshua Herman and Aleksei Sorokin for discussion. We also thank Karan Bakshi and Sahil Malhotra for their contribution to the repository.

We are grateful for the opportunities of presenting IMFDataPy in the conferences **PyData Global 2022**; **Midwest Machine Learning Symposium 2023**; and **The 15th R/Finance Conference for Applied Finance Using R**.