IMF Data Discovery & Extraction

Irina Klein

Master of Data Science Program
College of Computing
Illinois Institute of Technology

Joint work with Dr. Sou-Cheng T. Choi, Illinois Tech and SAS Institute Inc.

PyData Global 2022

December 3, 2022

Available Data

Dataset, a.k.a. 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
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

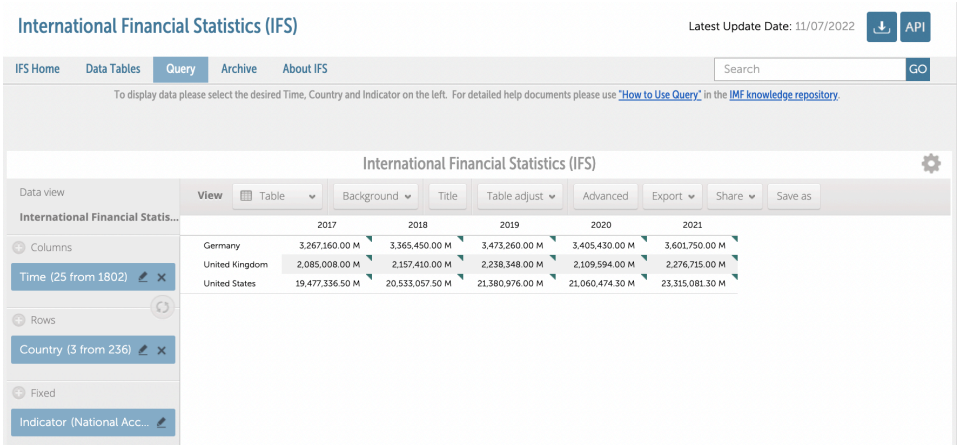
Table 1: Partial list of >30 monthly, quarterly, or annual data series for 190 member countries from 1960's to present available from IMF [IMFa].

Data Extraction Methods

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

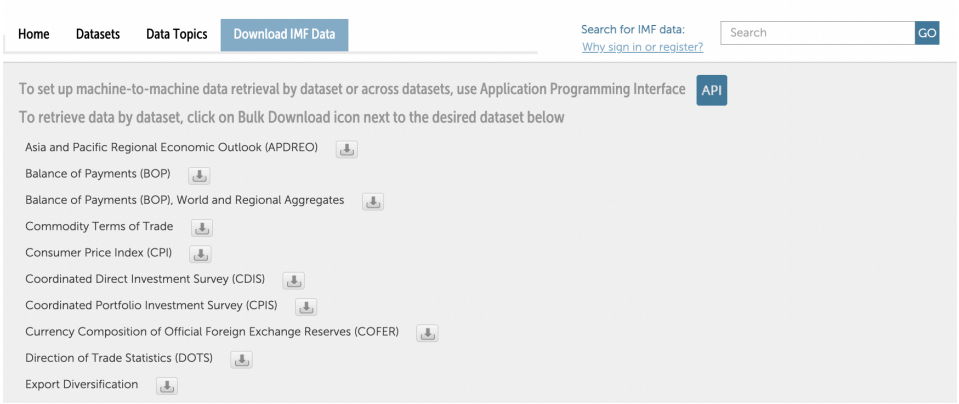
For detailed IMF help and documentation, see [IMFc].

Web Query Interface



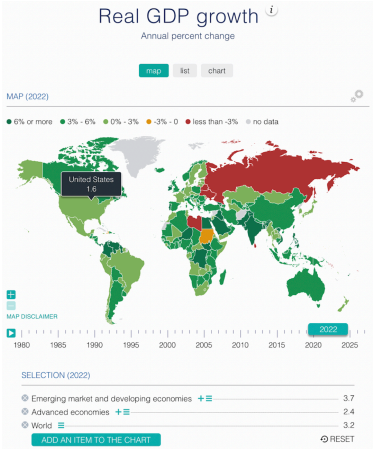
Real GDP over the last 5 years in the US, UK, and Germany [IMFd].

Bulk Download



Interface to bulk download an IMF dataset [IMFd].

IMF Data Mapper and IMF Mobile App



Real GDP trend accessed through a web browser or the IMF mobile app on a phone or a tablet [IMFb, IMF15]. On the left: Map view; on the right: Chart view.

JSON RESTful Web Service API

JSON

A file format for data storage and transmission that consists of key-value pairs and arrays.

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

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

RESTful API

An interface to interact with resources on a server using the **RE**presentational **S**tate **T**ransfer design pattern. The main building blocks are the request from the user and the response from the server [Tec].

Example 2. Examples of using the IMF RESTful JSON API: [Dew16, Dew]. Additional examples are available in the **imfdatapy** Github repository (in the folder 'demo').

IMF-Specific Definitions

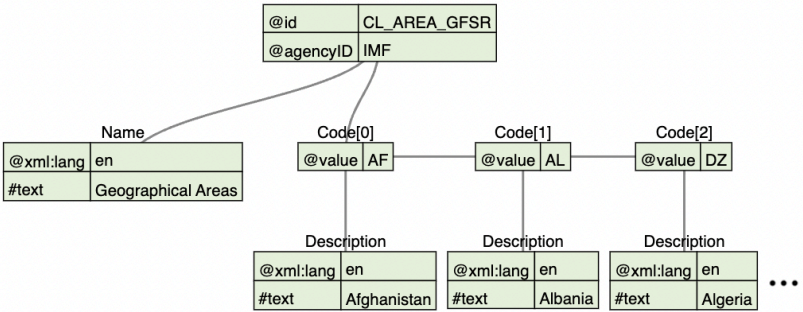
- **Series** is a dataset containing economic indices.
- **Indicator** is a set of time-indexed numeric values that represents an economic index or metric.
- **Dimension** is metadata for all indicators in an IMF data series. For example, in IMF series, most commonly the dimension are as follows:
  - Area[1] (e.g., the US)
  - Frequency (e.g., Quarterly)
  - Period (e.g., from 2020 to 2022)

[1] Area (or in some datasets 'Country') does not in all cases refer to a territorial entity of a state understood by international law. It may refer to a territorial group, such as the Eurozone or non-sovereign territories for which statistics are maintained.

Other Useful Definitions

- **Metadata** is a set of data that describes and gives information about other data.
- **Python package** is a collection of installable, reusable Python code modules that performs specified tasks.
- **Dataframe** is a two-dimensional tabular data format with columns of various data types.
- **Pandas Dataframe** is a dataframe from the Python package, Pandas.

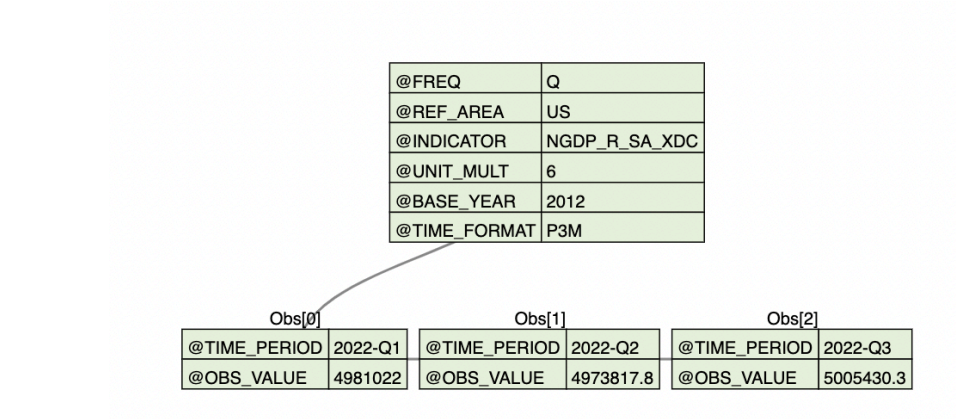
Metadata Structure



GFSR Area Dimension: first three items.

Note: This diagram is produced with JSON Tree Visualizer [Iva].

Data Structure



Quarterly Real Gross Domestic Product observations for the US in 2022.

Note: This diagram is produced with JSON Tree Visualizer [Iva].

Real GDP QoQ Change Example

Real Gross Domestic Product (Real GDP) is the inflation-adjusted monetary value of goods and services in a country in a given period of time.

Quarter-on-quarter (QoQ) is a change in performance between one quarter and the previous quarter.

To access the data using the JSON RESTful API, we need to:

- understand / lookup the dimensions of the series (index codes, country codes), using the DataStructure method[2], and specify the necessary dimensions
- request the data using the CompactData method

Python Notebook Example

[2] Description of the available methods can be found in GitHub repository and on IMF website [IMFe]

Problem

- Structure of the JSON datasets varies from series to series. It is too complex for an unprepared or non-technical user and requires time to understand the dimensions and the data available to be able to successfully load the data.

Existing work

Table 2: Existing packages to explore and extract the IMF data.

Name	Language	Functionality	Available datasets	Limitations
IMFData [Lee16]	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.
datapungi_imf [Ott20]	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.
imfpy [Kea21]	Python	Search through the datasets, download data into pandas dataframe, visualize data	DOTS	Only one IMF dataset.
PyIMF [Egg22]	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 (working) [CK22]	Python	Search through series names. Download data & metadata given series name and index **search terms** into Pandas dataframes and csv files	BOP, DOT, FSI, GFSR, IFS	Current limitations are described as part of the future work at the end of the slides

1. International Monetary Fund (IMF) data
2. JSON RESTful Web Service API
3. Proposed solution: **imfdatapy** package

Python **imfdatapy** package

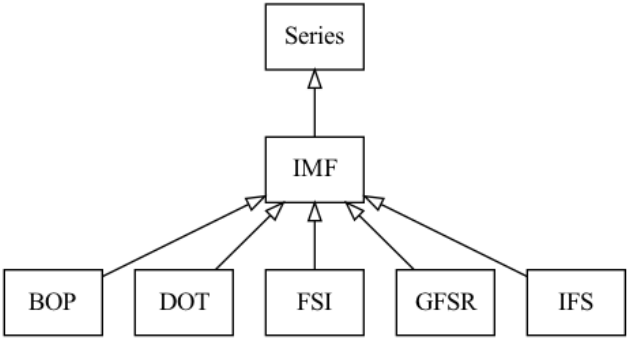
We propose creating an API wrapper package in Python to

- 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
- provide an intuitive way to search through the series or indicator names
- create detailed, searchable documentation
- ensure functionality with unit tests.

The current work is released as a Python package [BT22] and can be found at the GitHub repository:
https://github.com/Economic-and-Financial-Data-Discovery/imfdatapy.

imfdatapy package architecture

The package is designed in an object-oriented way, with children objects inheriting functionality from 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**.



Current classes in `imfdatapy`

Real GDP QoQ Change Example Using `imfdatapy`

To install the package run the command:
pip install imfdatapy

Now, the same example using the **imfdatapy** package.
Python Notebook Example.

Conclusions and Future Work

We have designed an extensible software architecture with simple APIs in the Python package, **imfdatapy**, for discovering and extracting IMF data series, pending the following tasks:

- Create comprehensive, searchable documentation
- Implement more robust input parsing
- Add all the IMF series with one or more unit tests for each series
- Perform large-scale testing
- Provide summary / descriptive statistics of the data and improve metadata visualization
- Collect and create more economic and financial use cases using **imfdatapy**

References I

[BT22] Tomas Beuzen and Tiffany Timbers, *Python packages*, CRC Press, 2022.

[CK22] Sou-Cheng T. Choi and Irina Klein, *imfdatapy*, https://pypi.org/project/imfdatapy/, 2022.

[Dew] Brian Dew, *IMF API with Python: An example*, https://www.bd-econ.com/imfapi1.html, Accessed: 2022-11-26.

[Dew16] ———, *Machine reading IMF data: Data retrieval with Python*.

[Egg22] Carlos Eggers, *Pyimf*, https://pypi.org/project/PyIMF/, 2022.

[IMFa] IMF, *Available datasets listed alphabetically*, https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sId=1479329132316, Accessed: 2022-11-26.

[IMFb] ———, *IMF Data Mapper*, https://www.imf.org/external/datamapper/, Accessed: 2022-11-26.

[IMFc] ———, *IMF help documents and references*, https://datahelp.imf.org/knowledgebase/topics/69748-help-documents-and-references, Accessed: 2022-11-26.

References II

[IMFd] _____, *International Financial Statistics - IMF data*, https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b, Accessed: 2022-11-26.

[IMFe] _____, *JSON RESTful web service*, https://datahelp.imf.org/knowledgebase/articles/667681-json-restful-web-service, Accessed: 2022-11-26.

[IMF15] _____, *IMF videos. IMF new mobile app available now!*, https://www.imf.org/en/Videos/view?vid=3984275247001, 2015.

[Iva] Ivan Ivanov, *Online JSON to tree diagram converter*, https://vanya.jp.net/vtree/, Accessed: 2022-11-26.

[Kea21] Liam Tay Kearney, *imfpy*, https://pypi.org/project/imfpy/, 2021.

[Lee16] Mingjer Lee, *Imfdata*, https://github.com/mingjerli/IMFData, 2016.

[Ott20] James Otterson, *datapungi_imf*, https://pypi.org/project/datapungi_imf/, 2020.

[Tec] IBM Technology, *What is a REST API?*, https://youtu.be/lsMQRaeKNDk, Accessed: 2022-11-26.

Thank you!

Contact Info:
Irina Klein, `iklein@hawk.iit.edu`
MAS Data Science, Illinois Tech

https://pypi.org/project/imfdatapy/
https://github.com/Economic-and-Financial-Data-Discovery/imfdatapy