

FAOSTAT 2.3

A revitalisation of the API wrapper of the FAOSTAT API

Sebastian Campbell

April 5, 2023

The FAOSTAT package is an important part of the Food and Agriculture Organization (FAO)’s image that is being maintained, but requires a makeover. Here an updated version, 2.3.0, of the package is presented with repaired access to FAO’s API, new functions and preparation for a complete overhaul in 3.0.0. The package has been modernised according to new coding conventions with improved dependencies, documentation and tests. Old useless functions have been pruned and the package is now firmly focused on providing an interface to FAO data to users of R.

Other formats

 [HTML report](#) |  [PDF report](#) |  [Presentation](#) |  [GitHub](#)

Project background

The motivation for this project came from a Data Mining project from UniLaSalle. It was suggested that we use FAO¹ data from their statistical platform FAOstat². As R was the language of choice, the obvious port of call was the FAOSTAT package³ (Kao, Gheri, and Gesmann 2022), developed by employees at FAO.

¹ Food and Agriculture Organization of the United Nations

² Food and Agriculture Organization Corporate Statistical Database

³ For the purposes of clarity, this document will use the style “FAOSTAT” for the R package and “FAOstat” for the statistical platform

However, the FAOSTAT package did not work. It could not download data from the API and could only download bulk data with the entirety of a dataset in one go. For the particular dataset we were interested in, we found that there was a discrepancy between the data in the bulk download and the data on the web platform.⁴

Eventually it became necessary to use the same API that the FAOstat website uses to pull data. This method worked and it became clear that it could be used to revitalise the FAOSTAT package and part of an effort to restore it to full functionality.

⁴ This discrepancy has been fixed as of 2023-03-10

FAOstat

FAOstat is FAO's web-based statistical platform for the free dissemination of food and agriculture statistics. This data is obtained from questionnaires that FAO distributes throughout the world every year (Food and Agriculture Organization of the United Nations 2019). Some of its data also comes from imputations and models where data is not available, but official country data takes precedence.

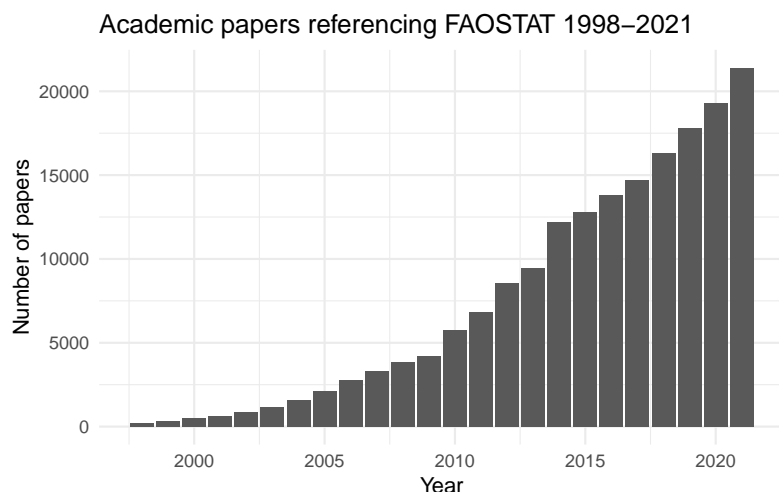


Figure 1: Academic papers referencing FAOSTAT over the last 25 years (Strobel 2018)

The FAOstat service is a public-facing aspect of FAO, with an overall trend of increasing citations in academic papers year on year with 21400 citations by 2021 (Figure 1).

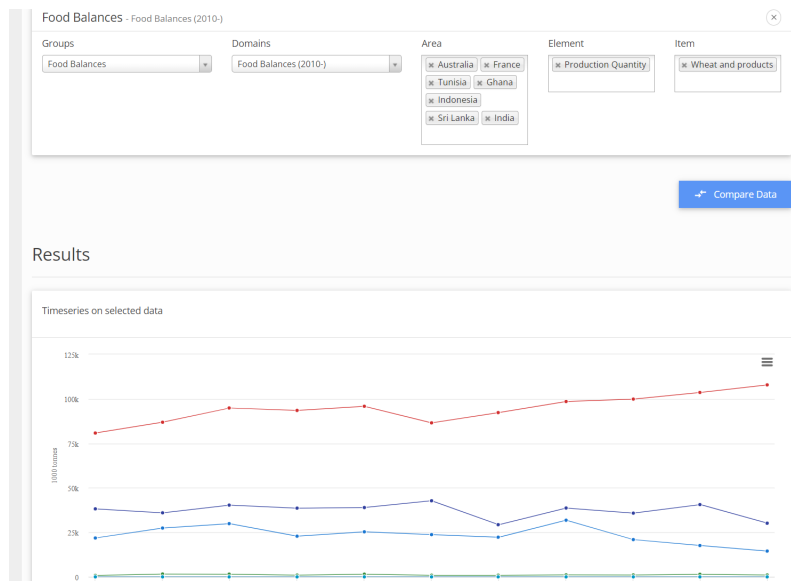


Figure 2: FAOstat interface for exploration of country data

This platform uses a REST API internally to communicate with its database as well as providing a set of zip files with the entirety of certain datasets in order to reduce the load on the database. This REST API allows the website to generate CSVs as well as to allow exploration of the data via interactive graphs (Figure 2).

FAOSTAT package

The FAOSTAT package is an API wrapper to pull data from FAO-STAT into a R session. It can also perform small necessary tasks such as country code conversion and coalescing data from different country groups.⁵

History

The FAOSTAT package was originally developed in 2013 as a tool to source data for the SYB⁶ project. The yearbooks are yearly summaries of the worldwide state of agriculture for that year. At the time, they were manually typeset and compiled. The new SYB project was to use a combination of LaTeX, knitr and R to automatically pull data from FAOSTAT and other data sources such

⁵ For example, China may be just the mainland or may include Taiwan (Chinese Taipei), Hong Kong and Macao

⁶ Statistical Year Book

as the World Bank. This data would be then be transformed and processed to create graphs and tables before finally formatting and typesetting to create a finished product which could then be printed.⁷ Given that this use case no longer exists, the primary use of this package is for researchers and other R users to read data from FAOstat in a clean way that makes it easier to move to analysis afterwards.

⁷ The author has no insight into the current production of the SYB, but they are still being produced and can be found on the [FAO website](#)

It is a reasonably popular package in the 86th percentile of all packages on CRAN on 2023-04-01 by downloads. In total, the package has been downloaded over 50 000 times with a peak 121 daily downloads on 2019-05-15. (Li 2023)

The package was maintained by Michael Kao, the author, from 2013 to 2014. In 2014, it was maintained by Filippo Gheri before passing to Paul Rougieux (the current maintainer) in 2020.

While it was [originally hosted on Github](#) under Michael Kao's personal account, it is [currently hosted on GitLab](#) under Paul Rougieux's personal account.

Current state

The FAOSTAT package has only a shadow of its former functionality. While it has retained the ability to download and process zip files and country code processing functions,⁸ its capacities are limited by the following issues:

⁸ For a full description of the status of individual issues, please see the GitLab issue [#20 Remove functions linked to defunct uses of FAOSTAT](#)

Functionality locked to the Statistical Yearbook

A number of functions are simply designed to pull in data from other sources such as the World Bank and to process that data into a format easily consumed by the Statistical Yearbook. As the yearbook no longer uses the FAOSTAT package, these functions have no further purpose, serving only to clog up the package and its help files.

Functionality powered by local files

Many uses of FAOSTAT require data outside of the data that comes directly from FAOSTAT. The major use case is for code conversions. There are two main code types that require conversion:

- Country codes
 - **FAO**: FAO's internal codes for countries⁹
 - **M49**: The UN standard country codes
 - **ISO2 & ISO3**: 2 and 3 letter country codes
- Item codes
 - **FAO**: Internal FAO codes to describe commodities
 - **CPC**: Central Product Classification code (United Nations Statistical Division 2005)

⁹ For further details about FAO and how it handles country identification, see [FAO's NOCS database](#)

The conversions are not dynamically taken from the API but rather stored in a fixed file. This makes them vulnerable to code or name changes in the future such as the name change of Swaziland to Eswatini in 2018 (United Nations 2018).

Change of FAOstat API

The FAOSTAT package is currently configured to access a now-defunct API¹⁰ (FAOSTAT3). As a result, it has no methods of retrieving data from the FAOstat platform with the sole exception of the bulk zip downloads which have been since adapted to use the current platform.

¹⁰ Originally hosted at faostat3.fao.org

Other issues

The FAOSTAT package is currently maintained by Paul Rougieux who has done an excellent job of keeping the package afloat. However, it's a small project done in his spare time, so he hasn't been able to make the time to do a full overhaul. In addition, he isn't an employee of FAO, but rather of the European Commission. As a face that FAO shows to the world, it seems reasonable that it be placed under its guidance.

As the package is old, it also has a number of dependencies that are unused or will be after redundant functions are removed. Other dependencies are simply no longer developed or have been superseded by newer packages.

Project goals

There are four main goals of this project:

- Fix up core functions
- Triage existing functions
- Characterise relevant aspects of the new API to wrap
- Transfer maintainership

Fix up core functions

The first priority to bring the FAOSTAT package to an acceptable level of functionality is to repair the most basic and most used functions.¹¹

Functions should not only be repaired but also renamed to have a consistent naming system.

¹¹ The functions below have been paired with a corresponding Gitlab issue that gives more context

Triage existing functions

As described in the [Current state](#) section, many functions no longer work and need to be removed. This involves a thorough listing of them and their functionality and assessment of their usefulness.

Characterise relevant aspects of the new API to wrap

Full overhauling the package is not in scope for FAOSTAT 2.3.0, but rather reserved for the next major version FAOSTAT 3.0.0.¹² This requires the additional step of inspecting the API manually as it doesn't have complete documentation.

¹² For a full set of changes expected for FAOSTAT 3.0.0, see the [GitLab milestone page](#)

Transfer maintainership

Uploading a package on CRAN¹³ requires the permission of the maintainer. At the upload step, an email is sent to the maintainer of the last recorded package version to confirm that the upload was authorised. Only then can the new version proceed to the registration step.

¹³ Comprehensive R Archive Network, the centralised repository where the majority of R users download R packages

Methods

On 2023-03-30, the FAOSTAT package version 2.3.0 was published to CRAN.¹⁴ The process required the resolution of the issues detailed in the [Project Goals](#) section.¹⁵

¹⁴ [CRAN package page](#)

¹⁵ Please see [GitLab Milestone 2.3.0](#) for a full list of related issues

In order to solve these issues, a full examination of the package context, the API and a redesign of existing functions needed to be made.

Learning package context

The two most important developers of the FAOSTAT package are the package author, Michael Kao and the current maintainer, Paul Rougieux.

Paul Rougieux has been very helpful in guiding the author's proposed changes to the package - making excellent suggestions in regards to code style and formatting as well as documentation and general useability.

An interview with the package author clarified that a number of of unusual functions were due to its previous linkage with the Statistical Yearbook and that the paper attached to the package had never been published and suggested that it should be tidied up and published in future.

Examining the API

The API is not fully documented and this is mostly because it's not really intended for interactive use. The main documentation is a JSON:API specification,¹⁶ which describes the structure perfectly, but gives no context or recommendations for use, nor does it describe output.

It was thus necessary to use this document to manually examine all the endpoints and characterise all the the data output from them.¹⁷

¹⁶ [JSON:API](#) is a standard way to describe the structure of an API as a single json document

¹⁷ A full set of this examination is found in [GitLab issue #23](#)

Redesigning functions

The existing functions to be refactored to read data from the new API. FAOSTAT 2.3.0 only has core functions in scope.¹⁸ The most core functions include:

- `getFAO`¹⁹ - The heart of the package, pulls a custom slice of data from FAOstat. Renamed to `read_fao`
- `FAOsearch`²⁰ - Allows a user to find the dataset they're looking for using the directory in FAOstat. Renamed to `search_fao`
- `translateCountryCode`²¹ - Translates country codes between formats. Renamed to `translate_countrycodes`

¹⁸ A full list of functions to be refactored or discarded is in [GitLab issue #20](#)

¹⁹ [GitLab issue #16](#)

²⁰ [GitLab issue #22](#)

²¹ [GitLab issue #28](#)

Consequent work

The above changes sometimes entailed additional work due to unforeseen problems or hidden requirements

Function regression

It was intended that the bulk zip download files remained intact, as they were the most functional part of the package. However, it depends on `FAOsearch/search_fao` which was changed as part of the 2.3.0 release. As a result, it was necessary to refactor some of the bulk download code to restore functionality.

Caching

One of the drawbacks of moving from internal data to pulling from the API is the need to read in data every time a lookup table or other piece of reference data is required. This slows down scripts as it's forced to wait for a reply from the FAOstat server. To mitigate that, caching was implemented.²² Now instead of pulling every time, the server is polled only once and the response is kept and referred to for subsequent queries.

²² [GitLab issue #27](#)

Metadata functions

When using the package, it was found that certain functionalities were immediately necessary. It was nigh-impossible to practically make a request for data without knowing what datasets were available and what their column names were. As a last minute addition, functions to perform these tasks were slipped into the 2.3.0 release.²³

²³ [GitLab commit: Added functions to get dimensions](#)

Testing

When making changes to a codebase it's important to make tests to assure that the function works as intended but also to report when the function breaks as a result of changes to it or its dependencies.

Making tests is repetitive work however and many tests are structured the same way. This sort of repetitive intellectual work is perfect for an AI. The author used chatGPT to help generate tests.²⁴ Gave it code and told it to give me ideas for tests

²⁴ [ChatGPTv3 conversation](#) 2023-03-28

Documentation

Documented using roxygen2 (Wickham et al. 2022), tests written using testthat (Wickham 2011)

- examples

Future work

- Release 3.0.0
 - Fully integrating with the new API
- Publishing a paper in [JOSS](#)
- Publish news in *rweekly*
- Publish FAO news alert
- Move package to FAO repository

Funding declaration

This project has been funded by the Food and Agriculture Organization of the United Nations and the author is grateful for their help in reviving it.

Food and Agriculture Organization of the United Nations. 2019. “Data collection.” https://www.fao.org/statistics/data-collection/en/#jfmulticontent_c728270-2.

Kao, Michael C J, Filippo Gheri, and Markus Gesmann. 2022. “FAOSTAT: Download Data from the FAOSTAT Database.” <https://gitlab.com/paulrougieux/faostatpackage> <https://cran.r-project.org/package=FAOSTAT>.

Li, Peter. 2023. *packageRank: Computation and Visualization of Package Download Counts and Percentiles*. <https://cran.r-project.org/package=packageRank>.

Strobel, Volker. 2018. “Pold87/academic-keyword-occurrence: First release,” April. <https://doi.org/10.5281/zenodo.1218409>.

United Nations. 2018. “Eswatini.” United Nations. <https://www.un.org/en/about-us/member-states/eswatini>.

United Nations Statistical Division. 2005. “Central Product Classification (CPC) Version 2.1.” Department of Economic; Social Affairs, United Nations. [https://unstats.un.org/unsd/classifications/Econ/Download/InText/CPCv2.1_complete\(PDF\)_English.pdf](https://unstats.un.org/unsd/classifications/Econ/Download/InText/CPCv2.1_complete(PDF)_English.pdf).

Wickham, Hadley. 2011. “testthat: Get Started with Testing.” *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham, Hadley, Peter Danenberg, Gábor Csárdi, and Manuel Eugster. 2022. *roxygen2: In-Line Documentation for R*. <https://cran.r-project.org/package=roxygen2>.