

Utilizing our external APIs

Jan Bruusgaard, Statistics Norway, jbr@ssb.no

Historical context

In the early days of the World Wide Web (www) it was common for the National Statistical Institutes (NSIs) to provide only a limited number of output tables for free (typical main aggregations of popular topics). The systems connected to the output databases were paid services where users paid per minute they were connected. Very few users wanted to pay for the service and the user interface was complicated. Hence, the establishment of Statbanks as a free service some 20 years ago was a huge improvement.

Around the same time Denmark, Sweden and Norway agreed on a Common Nordic Metadata Model (CNMM), which is still in use. The Nordic countries also agreed on development and using parts of the PC-Axis family of products. The web evolved, and it became important to integrate the web with other software and services. Excel got “Web queries” and for the StatBank it became important to develop *web services* for exchange of data and metadata.

In 2013 Statistics Norway released our first version of an API (Application Programming Interface) open for external users. This API is very limited, based on the most frequent used tables from our Statbank. At the same time the Px-family product PxWeb got an API, the PxWebApi. This made it possible to program queries for extracting data from PxWeb. Both Statistics Norway and Statistics Sweden decided now to use PxWeb themselves as the Statbank frontend. In May 2016 Statistics Norway launched their version of the PxWebApi.

Today all Nordic NSIs offer an API towards their Statbanks, including Greenland, Faroe Islands and Åland. Except for Denmark, all are based on the software PxWeb. As Denmark use the same data/metadata model, CNMM, their API share many of the features. This common Nordic platform gives us many benefits.

What is an API?

API (Application Programming Interface) makes software talk to software, often described as machine-to-machine communication. Open web APIs are accessible for any machine connected to internet. But over time the human aspect has been given more and more weight, and in 2021 Gartner group gave this definition:

“APIs enable applications to connect to other applications and to data, as well as engage developers and provide the foundations of business ecosystems.”

Most international organisations also offer APIs for statistics, such as Eurostat, UN, World Bank and IMF.

Statistics Norway offers another API for Statistical Classifications and Codelists (Klass), where you can get standards and correspondence tables, and track changes.

The main use of our PxWebApi is that a user wants to define a query to get updated figures. These can be shown as a table or chart on a webpage or imported to software like R, Excel, PowerBI, SAS or Jupyter Notebooks for further processing.

Open source

Open source makes IT development work more like research where it is possible to build on others' work. It is possible to make new building blocks or modify and improve these blocks. The blocks are often called packages or libraries. This ecosystem has made the utilization of our APIs much easier.

For getting data via NSIs APIs there are Open-source tools available, see Annex. Quite a few builds on each other. When you have a general tool for processing the output, a tool built on top can simplify the programming for accessing our statistics. This ecosystem of open source prevents one from reinventing the wheel. An important task for an NSI is to make these tools known. Statistics Netherland does an important job by maintaining the list “Awesome official statistics software” (<https://github.com/SNStatComp/awesome-official-statistics-software>).

The main place for publishing Open source is github.com. Github is owned by Microsoft. PxWeb itself became Open source in 2021, <https://github.com/statisticssweden/PxWeb/>.

Who are the users?

It is easy to think that IT developers would be the main user group of an API, but this is only partly true. In my experience few developers on their own initiative start to deliver a service towards our API, unless they can develop or integrate our data in a commercial product.

Just as often the API users are professionals as researchers in other subjects, e.g., medicine, or economists with IT knowledge. If the organisation is large or have financial resources, they can engage their own developers or external consultants to develop what they want. This is the case for the central bank and companies within finance sector, Central government authorities within environment protection and education who rely on massive amounts of statistics from us. Regional government has less resources and do more themselves. Also, within an NSI reading and writing code has been a task for employees outside the IT department. This is also the case in higher education where exercises and examples use data via our API in the fields of economics and political science.

There is a substantial difference among users and the type of statistics they want. A special kind of users are those interested in figures as *fast as possible* after release. This is the case when we release CPI, PPI and National accounts. These days there is continuously heavy traffic. To reduce the pressure these days it is important to play on team with the mediators like Bloomberg, Refinitiv, Macrobond and Reuters. We should not see these media as competitors or troublemakers. They do an important relief job.

Their use our APIs is as a form of data assisted journalism. When Bloomberg just after release of our foreign trade data publish one-liners like this on their terminals:

2022M03 Not - 138 371
NORWAY MARCH TRADE SURPLUS NK138.4 BLN

They make a combination of a template (green) and our data (callout) and some processing in Python to transform our data into the form Bloomberg want to present as news to the financial markets.

This shows that the use of our APIs in journalism is most appropriate when there can be some kind of repetition. For large projects in investigative journalism, their use of our data is more like a one-time job. Then the use of our APIs is not relevant.

Many services use the API to integrate our data with other services, but only a few are public.

For NSIs own use the API is used for building pages with charts, tables and dashboards that updates themselves when there are new data. It can be used also for automatic checks in the publishing process.

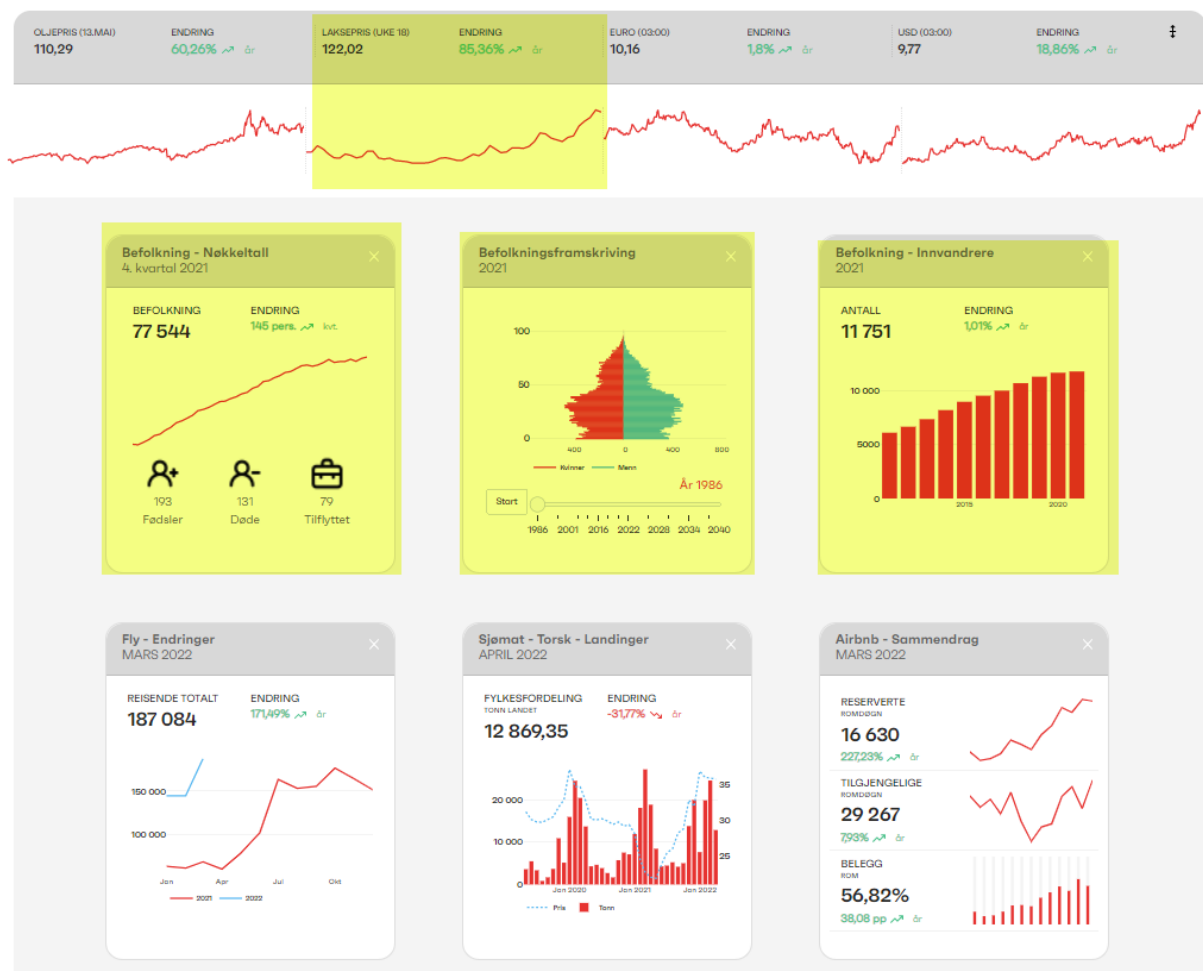
The transformation at Statistics Norway towards a new data platform (DAPLA) makes more use of Open-source tools like Python and R packages combined with Jupyter notebooks. Our external APIs has also been used for training in these new tools, both PxWebApi and the Classifications and codelists API.

Products based on Statistics Norway's APIs

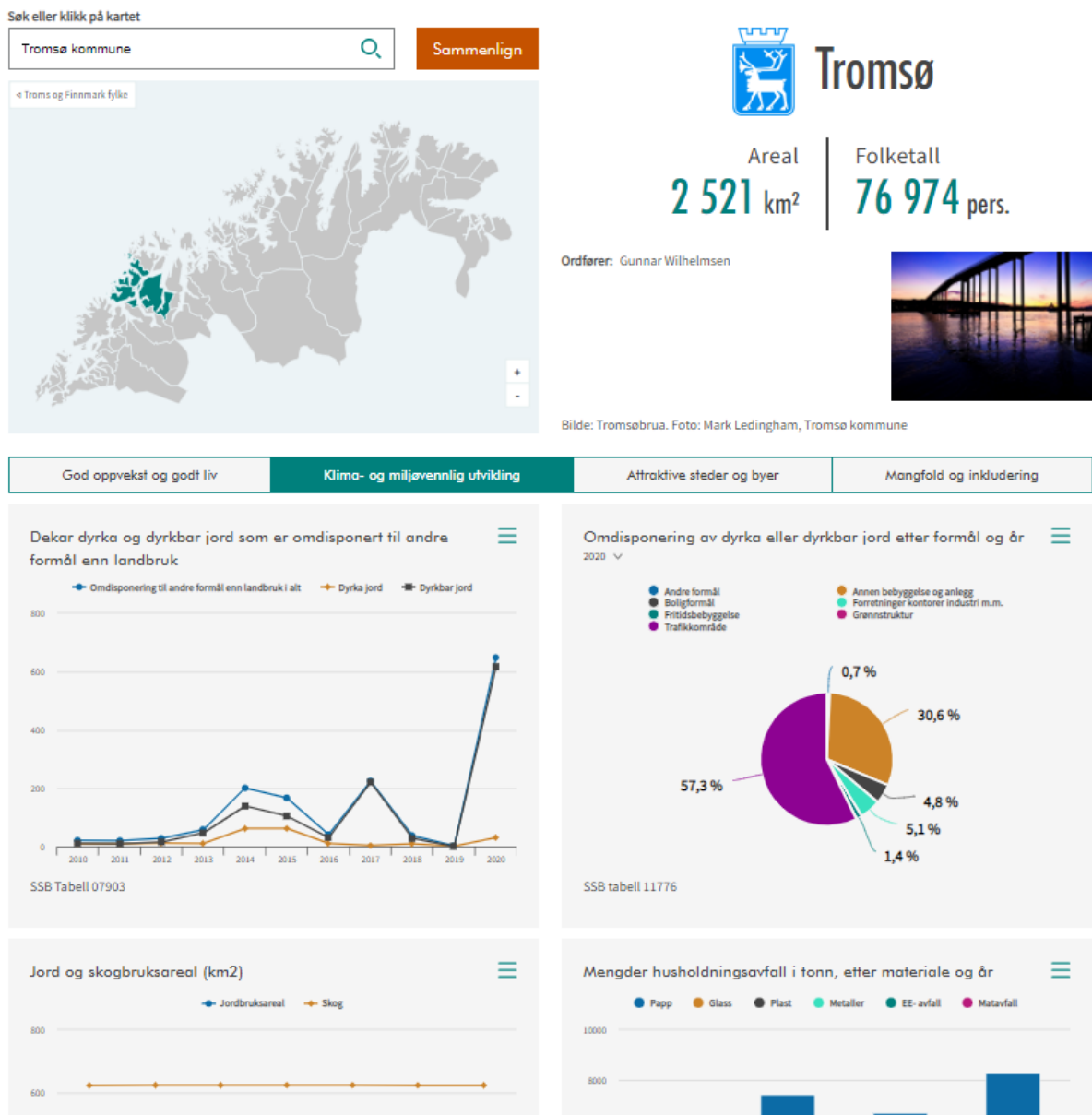
It is hard to track the products that use our API because they often are commercial products. Some are public like "**Tellern**" from Kunnskapsbanken Nord-Norge (<https://telleren.kbnn.no>). This web application is made in R + Shiny.

Tellern shows our regional statistics combined with other sources like Central bank data, Airbnb etc.

Extract of page from Tellern for the municipality Tromsø, SSB figures in yellow.



The **Norwegian Association of Local and Regional Authorities (KS)** have made their own service for regional statistics mostly based on data via our APIs.



<https://www.ks.no/status-kommune/kart/region?r=5401>

The company **Framsikt** (<https://framsikt.no>) have more than half of Norway's municipalities as their clients. They use a huge amount of our data via our API and combine these with other data mainly from the municipality itself for their business management.

ERC (Economic RESTful Client) (<https://zhnzhn.github.io>), gives the ability to explore and compose data from 29 data providers to charts. Here you can get data from Nordic APIs, Eurostat, UN, Quandl and DBnomics.

ERC (Economic RESTful Client) is a web app that gives the ability to explore, visualize and compose economic and financial data mostly to charts from open and private data providers.

▲ Data Providers (All 29):

[Quandl](#) (50) [DB NOMICS](#) [Eurostat](#) [UN Comtrade](#) [World Bank](#)

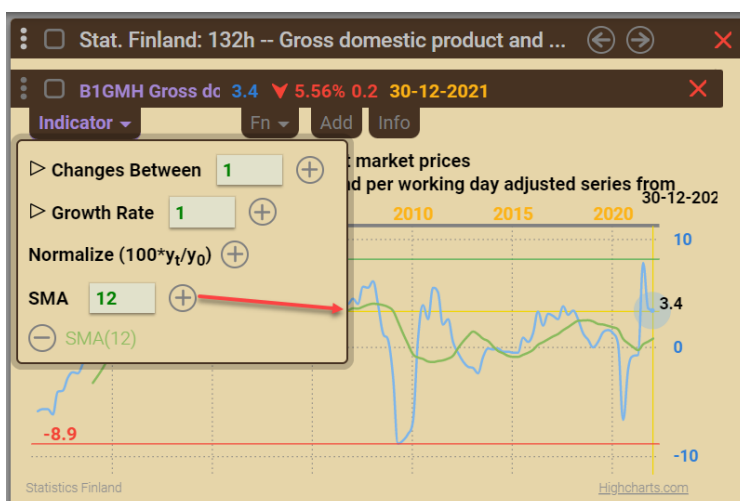
Insee: [France Statistics](#) [ONS: UK Statistics](#) [Statistics Norway](#) [Statistics Sweden](#)

[Statistics Finland](#) [Statistics Denmark](#) [CSO Ireland](#)

[U.S. Bureau of Labor Statistics](#) (25) [CryptoCompare](#) [CoinGecko](#) [CoinMetrics](#)

[CoinLore](#) [Coinpaprika](#) [Binance](#) [Bitstamp](#)

It is quite advanced, directed for economic professionals, but the user interface and design have room for improvement.



Q4opendata <https://stat.qumio.com> is a service for generating API query code for Excel and PowerBI. originally made for the Finland's official TAX authority statistics, but has been extended to cover all PxWebApis

Q4OpenData - Query for OpenData

There was a requirement for a name so here it is! This Service is now being called as Q4C

The screenshot shows a scrollable list of data sources. The categories and their sub-items are:

- Statistics Finland Official stats: Explorer database
- Natural resources institute Finland Statistics
- Finnish Transport Safety Agency Trafi (beta)
- Official Stats Finland, Swedish**
 - Official Tax Authority Statistics, Swedish
 - Finnish Trafiks kerhetsverket Trafi (beta)
- Official Stats Sweden**
 - Statistics Sweden: Official Statistics, English
 - Statistics Sweden: Official Statistics, Swedish
- Official Stats Eurostats**
 - Eurostat, Finnish (Tilastokeskuksen eurotilastot)
 - Eurostat, English (Statistics Finland, European Statistics)
- Official Stats Norway**
 - Statistics Norway: Official Statistics, English
 - Statistics Norway: Official Statistics, Norwegian
- Official Stats Moldova**
 - Official Statistics Moldova: Economics, English
 - Official Statistics Moldova: Environment, English
 - Official Statistics Moldova: Population and demographic processes, English
 - Official Statistics Moldova: Social, English

On the right side of the interface, there is a 'PXWeb' logo and some partially visible text: 'BI / Excel r', 'available. I hav', 'is tool prov', and a URL 'www.qumio.c'.

Help and training

To inform and increase the use of the StatBank APIs we have done several measures

- Help and resource pages.
- Code examples in R, Python, JavaScript, also as Jupyter notebooks.
- Half day courses (physical and online) 3-8 times a year
- Short and long videos
- Participation user meetings, seminars etc.
- User support by statistikbanken@ssb.no
- Mailing list – mail a few times a year about news and big changes to the tables
- Social media (Twitter, LinkedIn)
- Hacks – national and at Statistics Norway

As it is hard to know who the users are, the half day courses have been an important source for knowing who the users are, and their intentions for use of the API.

A mailbox for user support is important for feedback. Most users do not get in contact with us unless something is wrong. Several times they have reported errors before we spotted them ourselves. Often these users who has got problems in addition add positive feedback about the API.

API pitfalls and challenges

An API is a vulnerable professional environment, and person dependent both with the API user and with the API provider (the NSI). The API users want stability and no changes. When the output is working, they want to keep it like it is for the future. Quite often we tell the owner of StatBank table that we would not recommend a change because it will break an API query or saved query towards their table. Some changes can be out of our control, like the Norwegian regional reform, where municipalities were merged and split. Changes in schema or variables often result in discontinued tables. A new table is made instead and the user have to modify their API-query. It is important that the API user maintain the competence to handle these changes. To set up API queries, accuracy is critical.

Statistics

The use of the API is possible to track, but the figures can be hard to analyse. Over the years the number of requests have gone a bit down, but the total amount of downloaded data cells has increased much more. In my opinion the figures “without cache” are the best for comparing API-figures over time.

Use of Statistics Norway's PxWebApi, without cache

	Number of downloads	Downloaded data cells	Cells pr request
2020	2 486 237	40 559 761 048	16 314
2021	2 638 398	62 208 113 961	23 578

Foreign trade by HS is the most frequent used table. But these are also the largest tables in our Statbank.

Output formats from the API

PxWebApi offers many different formats. This can be both a strength and a weakness. It gives a high grade of flexibility, but it also makes it hard to concentrate limited resources on improvements. Unless you want to invent the wheel yourself,

there are a quite confusing jungle of tools that can be used to utilise the output from the APIs.

The main output formats from PxWebApi are:

- CSV (comma separated values) in different flavours
- JSON-Stat
- SCB-JSON
- Px – Pc-Axis file
- Excel

Each NSI can decide which formats they want to support, and make as their default or recommended format, but all formats are accessible.

JSON (JavaScript Object Notation) is a lightweight data interchange format used all over the web. JSON-stat is subset and a modern format for effective representation for multidimensional data cubes i.e., statistical tables. It is also flexible for representing and include metadata.

Since 2013 Statistics Norway has tried to promote JSON-stat as the main format for our statistical APIs. This format is also supported by the APIs of Eurostat and World Bank. There has been made several tools to utilise JSON-stat. One challenge is that JSON-stat at first glance can be difficult to understand and have a steep learning curve.

At the same time Statistics Sweden (SCB) developed SCB-JSON as an output format which is more limited but easier to understand. This is the format SCB promotes on their API pages.

The Px-Axis format is a thoughtful, but old, compact text file format with very good metadata support. Px-files is the foundation for most PxWeb installations today. The Px file format was developed by SCB in the 1990s with a major revision in 2008. Statistics Finland is responsible for the tools to produce Px-files, Px-Map and Px-job.

At the same time CSV, with its well-known table layout, and Excel, are the formats known to all users, and easily accessible.

(What about SDMX maybe someone ask? That is an output format which is so complicated, that only a few wants to support it, mainly Eurostat and central banks.)

The result is that except for some minor improvements in JSON-stat and the CSV output in 2019, there has been no improvements in the output formats since the start of PxWebApi in 2014.

In my opinion the best direction for the output formats has been and should be to concentrate on developing JSON-stat to its full potential with at least the same metadata support as the old-fashioned Px file format. If this was done, the developers and users would see the benefits of the format and it would be easier to develop even better solutions on top.

Open data – open license

The definition of Open data is data published under an open license, which gives the right of reuse.

Most NSIs now use the license Creative Commons 4 BY, which gives free use if you give credit to the NSI as source. This can be a challenge for automation when combining our data with others, especially when using API. Wikipedia in Norway and Sweden are among the ones who have complained about this. US Census and other American agencies have for a long time published their statistics as public domain without the obligation to give source. A step in the right direction is that SCB now have changed their license to CC0 (<https://www.scb.se/en/services/open-data-api/>), which means that users are not obliged to give the NSI as source. It will be interesting to see if other nordic NSIs will follow.

Lessons learnt

The responsibility of development of the three external APIs at Statistics Norway are placed at different places in the organisation and all of them was released as kind of an MVP (Minimum Viable Product). After the launch of PxWebApi in 2016 it has been hard to get resources for more than better handling of the extreme pressure on our servers the first minutes after release of new data and bug fixes.

It has been very hard to get further functional improvements in the APIs. The result is that many important external users have to spend extra time and resources in order to get the data they want. In PxWebApi itself there is also lack of central metadata

(aggregations and groupings) which make it hard to query the API for the same data as from the web interface.

Neither SSB nor SCB have so far prioritised to spend resources on developing functional API improvements based on external user needs. This has led to other government agencies in Norway have spent a lot of resources on consultants and developers to overcome the limitations in our APIs.

The good news is that in April 2022 SSB and SCB started the common work on developing a new PxWebApi 2.0, which is planned to live side by side with today's version. This will improve the way you can query. the PxWebApi, as users can construct URLs for queries (http GET). It will be better listing of metadata, as aggregations and value sets which has been wanted by many. The API output formats themselves are currently not in the scope of this work, but there will be improvements here as well.

Appendix. An overview of the tools for utilizing Nordic statistical APIs

For extended information see separate Excel file: api-tools.xlsx.

Name / link	Description
Rjstat	Handle 'JSON-stat' Format in R
PxWebApiData	Function to read PX-Web data into R via API. Built on top of rjstat
JSON-stat Toolkit	A set of tools for utilizing JSON-stat, made by Badosa the creator of JSON-stat
Pyjstat	Library to handle JSON-stat data in python using Pandas DataFrames.
jsonstat.py	Another tool for getting JSON-stat dataset to jupyter notebooks. Older than pyjstat
Tablea WDC	JSON-stat Web Data Connector (WDC) for Tableau
json-stat.java	Implementation of JSON Stat in Java
JSONStat.jl	JSON-stat to Julia language
Stats-to-pandas	Python interface to get PxWebApi data to Pandas dataframe. Built on top of pyjstat
SSB API helper	Built on top of pyjstat
PxWebApiData call creator	Tool for converting PxWeb API queries into PxWebApiData format in R
PxR	Set of functions for reading PC-Axis files
pyaxis	PC-Axis statistical data format to Python Pandas dataframe converter
PxWeb	R Interface to PXWEB APIs
pyscbwrapper	Python wrapper for Statistics Sweden's API
Q4OpenData	Query for OpenData. Service generates Power Query (M) script from PxWebApis to PowerBI/Excel.

Name / link	Description
Awesome official statistics software	An awesome list of statistical software for creating and accessing official statistics
ERC (Economic RESTful Client)	Explore, visualize and compose economic and financial data mostly to charts from PxWebApis and others
Eurostat	PowerBI custom connector for Eurostat API
Eurostat API	Eurostat API - JSON-stat output. (The alternative output format is SDMX)
Eurostat's Swiss Army Knife	- A single solution for all your Eurostat data needs
KlassR	Functions to search, retrieve and apply classifications and codelists using Statistics Norway's Klass API

[World bank indicators API v2](#)

World Bank API support JSON-stat output

Name / link	Description
SSB Code examples	API code examples in Python, R, JavaScript and Powerquery (Excel/PowerBI)
JSON-stat Explorer	Useful tool for understanding JSON-stat
Power BI ESTP course	ESTP-webinar on Power BI held on the 20.-22. april 2021