TALL SOM FORTELLER

PX-API and JSON-stat
present state and future directions

JAN BRUUSGAARD, PX-MEETING 2018

Statistisk sentralbyrå
Statistics Norway

Agenda

PX need need a better branding

PX-API and JSON-stat news

Future directions and priorities in API

New API-package for R PxWebApiData

PX need some branding

- PX-web or Pxweb or PxWeb
- Google don't like hyphens
  - Two separate words
- PX-API or PxApi or px-web API or pxweb API or PX API
- Statbank or Statistical database

How can users know that we have the same API?

- Result: The wheel is reinvented

New in PX-API

- CORS support
- Improved performance
- Use of table id/table now from root URL of API
  - id: "SysselsYrkSek" -> http://data.ssb.no/api/v0/en/table/SysselsYrkSek
  - Can be used together with e.g. search for all tables
    - http://data.ssb.no/api/v0/en/table/?query=*&filter=*
- Also some improvements in the Norwegian console

Some PX-API tips

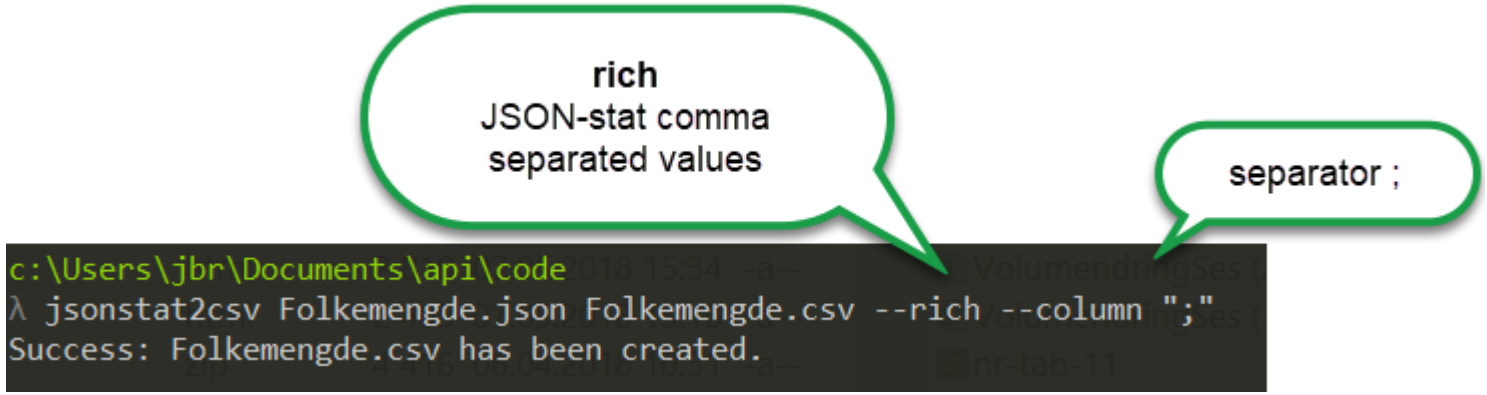
- Saved query as JSON-stat can work as GET url
  - Can be used together with libraries and tools
- Search for date, use: filter=published
  - Search tables published a specific date.

    https://data.ssb.no/api/v0/en/table/?query=20180507*&filter=published
  - Tables published in october

    http://data.ssb.no/api/v0/en/table/?query=201810*&filter=published

JSON-stat improvements

- JSON-stat is now published by 27 agencies
- Improvements in utilities suite and conversion
- New format: JSON-stat comma separated values (JSV)
  - Jsonstat2csv -- rich



```
jsonstat,.,|
label,"Population and population changes, by region, contents and year"
source,Statistics Norway
updated,2018-10-22T09:46:00Z
dimension,Region,region,1,0,The whole country,geo
dimension,ContentsCode,contents,4,Folkemengde,Population,Levende,Live birt
dimension,Tid,year,6,2013,2013,2014,2014,2015,2015,2016,2016,2017,2017,201
data
Region,ContentsCode,Tid,value
0,Folkemengde,201
0,Folkemengde,201
0,Folkemengde,201
0,Folkemengde,201
```

jsonstat,.,|
label,"Population and population changes, by region, contents and year"
source,Statistics Norway
updated,2018-10-22T09:46:00Z
dimension,Region,region,1,0,The whole country,geo
dimension,ContentsCode,contents,4,Folkemengde,Population,Levende,Live births,Dode,Deaths,Nettoinnflytting,Net migration,metric,0|,0|,0|,0|
dimension,Tid,year,6,2013,2013,2014,2014,2015,2015,2016,2016,2017,2017,2018,2018,time
data
Region,ContentsCode,Tid,value
0,Folkemengde,2013,5051275
0,Folkemengde,2014,5109056
0,Folkemengde,2015,5165802
0,Folkemengde,2016,5213985
0,Folkemengde,2017,5258317
0,Folkemengde,2018,5295619
0,Levende,2013,58995
0,Levende,2014,59084
0,Levende,2015,59058
0,Levende,2016,58890
0,Levende,2017,56633
0,Levende,2018,0

JSON-stat conversion, news

- Jsonstat2jsonstat – convert from JSON-stat version 1 to version 2
- Sdmx2jsonstat
  - converts sdmx-json (e.g. from OECD) to json-stat

More JSON-stat tips and news

- R combined with JSON-stat and package rjstat
  - The best way to use PX-API from SAS and PowerBI, as these programs now reads R code
- R package pxR (for PX-files) can convert px-files to json-stat
- *Statstransfer* conversion software now reads json-stat

New to JSON-stat, try: https://json-stat.com/explorer/

PX-API - Priorities for the future

My opinion:

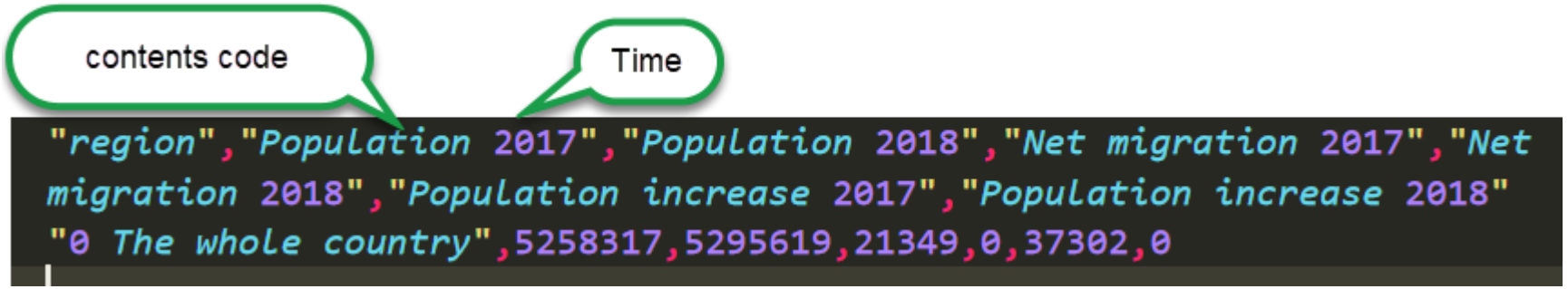
1. Improve existing formats
2. Improve queries
3. Support for JSON-stat v. 2.0
4. PX-API v.2

1. Improve existing formats

- JSON-stat 1.2
  - Published date
  - Include Datanotecell *
  - Footnotes
  - Include link to POST-url
  - Reduce entry level - logical (not random) order of elements

Improve today's formats

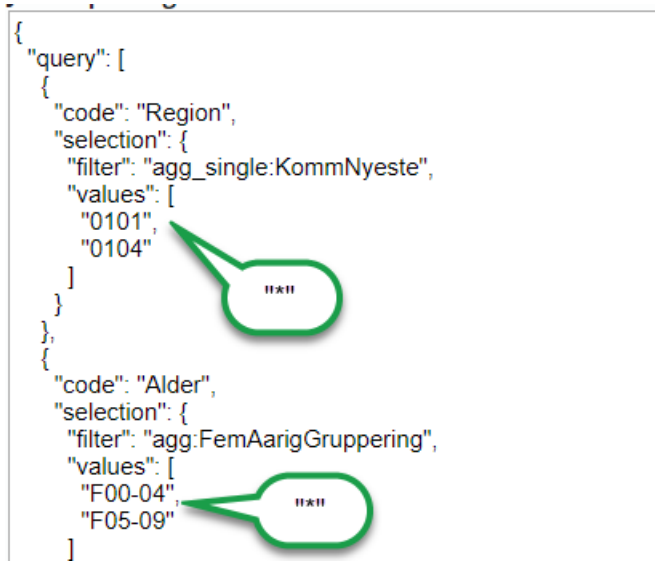
- CSV2 – move Contents code to first column



- CSV3 – with codes instead of text
  - no ÆÅØ etc makes CSV easier to import to other systems, e.g. SAS

Improve queries

- Possibility to mask one or more single char by using: **?**
- Possibility to use ***** in agg



Statistisk sentralbyrå
Statistics Norway

Sommeone?

- DCAT-AP (https://data.europa.eu)
- Documentation (Swagger / Raml)
- Third party packages becomes more important
  - Our role to stop reinventing the wheel?

R package PxWebApiData

Function to read PX-Web data into R via API

- Built on top of rjstat
- Made by Statistics Norway

- https://CRAN.R-project.org/package=PxWebApiData
- Works also with saved queries as JSON-stat
- Will also be included in RopenGov/pxweb package

Usage

```
ApiData(urlToData, ..., getDataByGET = FALSE, returnMetaData = FALSE,
  returnMetaValues = FALSE, returnMetaFrames = FALSE,
  returnApiQuery = FALSE, defaultJSONquery = c(1, -2, -1),
  verbosePrint = FALSE, use_factors = FALSE, urlType = "SSB")
```

Arguments

urlToData	url to data or id of SSB data
...	specification of JSON query for each variable
getDataByGET	When TRUE, readymade dataset by GET - works only for Statistics Norway readymade datasets
returnMetaData	When TRUE, metadata returned
returnMetaValues	When TRUE, values from metadata returned
returnMetaFrames	When TRUE, values and valueTexts from metadata returned as data frames
returnApiQuery	When TRUE, JSON query returned
defaultJSONquery	...ables not included in

To start in R

```
> install.packages("PxWebApiData")
```

```
> library(PxWebApiData)
```

```
> ?ApiData
```

- and follow examples

PxWebApiData

- Each variable is specified by using the variable name as input parameter.
- The value can be specified as: TRUE (all), FALSE (eliminated), imaginary value (top), variable indices, original variable id's (values) or variable labels (valueTexts).
- Variables not specified is set to the value of defaultJSONquery whose default means the first and the two last elements.

```
> #saved query as JSON-stat file
> x <- ApiData("https://www.ssb.no/en/statbank/sq/10012982/", getDataByGET = TRUE)
```

```
> x[[1]]  # The label version of the data set
                region      contents year   value
1  The whole country    Population 2014 5109056
2  The whole country    Population 2015 5165802
3  The whole country    Population 2016 5213985
4  The whole country    Population 2017 5258317
5  The whole country    Population 2018 5295619
6  The whole country   Live births 2014   59084
7  The whole country   Live births 2015   59058
8  The whole country   Live births 2016   58890
9  The whole country   Live births 2017   56633
10 The whole country   Live births 2018       0
11 The whole country         Deaths 2014   40394
12 The whole country         Deaths 2015   40727
13 The whole country         Deaths 2016   40726
14 The whole country         Deaths 2017   40774
15 The whole country         Deaths 2018       0
16 The whole country Net migration 2014   38155
17 The whole country Net migration 2015   29802
18 The whole country Net migration 2016   26076
19 The whole country Net migration 2017   21349
20 The whole country Net migration 2018       0
```

```
> x[[2]]  # The id version of the data set
   Region     ContentsCode  Tid   value
1       0      Folkemengde 2014 5109056
2       0      Folkemengde 2015 5165802
3       0      Folkemengde 2016 5213985
4       0      Folkemengde 2017 5258317
5       0      Folkemengde 2018 5295619
6       0          Levende 2014   59084
7       0          Levende 2015   59058
8       0          Levende 2016   58890
9       0          Levende 2017   56633
10      0          Levende 2018       0
11      0             Dode 2014   40394
12      0             Dode 2015   40727
13      0             Dode 2016   40726
14      0             Dode 2017   40774
15      0             Dode 2018       0
16      0 Nettoinnflytting 2014   38155
17      0 Nettoinnflytting 2015   29802
18      0 Nettoinnflytting 2016   26076
19      0 Nettoinnflytting 2017   21349
20      0 Nettoinnflytting 2018       0
```

TALL SOM FORTELLER

Comments?

Statistisk sentralbyrå
Statistics Norway