TALL
SOM FORTELLER

Uses of PxWebApi and PxWeb with focus on output formats

JAN BRUUSGAARD, STATISTICS NORWAY

PX-MEETING, ARMENIA 2019

Statistisk sentralbyrå
Statistics Norway

Agenda

PxWeb and PxWebApi formats – JSON-stat2

Norwegian specialities in PxWeb(Api) (video)

Uses of PxWeb API

How to use Relational table in Excel (video)

Further PxWeb Api development

Statistics Norway hackathon, Hack4SSB

JSON-stat version 2

- Better order of elements

JSON-stat

```
{
  "dataset": {
    "status": {
      "15": "..",
      "16": "..",
      "17": "..",
      "20": "..",
      "21": "..",
      "22": ".."
    },
    "dimension": {
      "ContentsCode": {
        "label": "contents",
        "category": {
          "index": {
            "Arbeidslause2": 0,
            "Arbeidslause4": 1,
            "Sysselsatt": 2,
            "Sysselsette2": 3,
            "Sysselsette3": 4
          },
          "label": {
            "Arbeidslause2": "Unemployed persons",
            "Arbeidslause4": "Unemployed persons in
```

Version 1

```
{
  "class": "dataset",
  "label": "08930: Employment and unemployment, seasonally
  "source": "Statistics Norway",
  "updated": "2019-10-24T06:00:00Z",
  "id": [
    "ContentsCode",
    "Tid"
  ],
  "size": [
    5,
    5
  ],
  "dimension": {
    "ContentsCode": {
      "label": "contents",
      "category": {
        "index": {
          "Arbeidslause2": 0,
          "Arbeidslause4": 1,
          "Sysselsatt": 2,
          "Sysselsette2": 3,
          "Sysselsette3": 4
        },
        "label": {
          "Arbeidslause2": "Unemployed persons",
          "Arbeidslause4": "Unemployed persons in per cent
```

Version 2

Statistisk sentralbyrå
Statistics Norway

JSON-stat2

- Coming in PX-web 2020, rel 1
- More user friendly
- Better prepared for the future
- Better and more modern serializer
  - no more 5 Mb size limit
- in line with Eurostat API

Other JSON-stat news

- Javascript
  - Badosa's tools now named as JSON-stat toolkit.
  - Support for ECMA script
    - Examples combining mulitple sources using promises
- R
  - We decided to keep Statistics Norway's PxWebApiData as a separate package and not be a part of the PxWeb package
- Python / Jupyter notebooks – work in progress
- JSON-stat to Excel / PowerBI first example

JSON-stat in PxWeb- local SN additions

- Updated – now showing published date,

  not time for retrieval of dataset

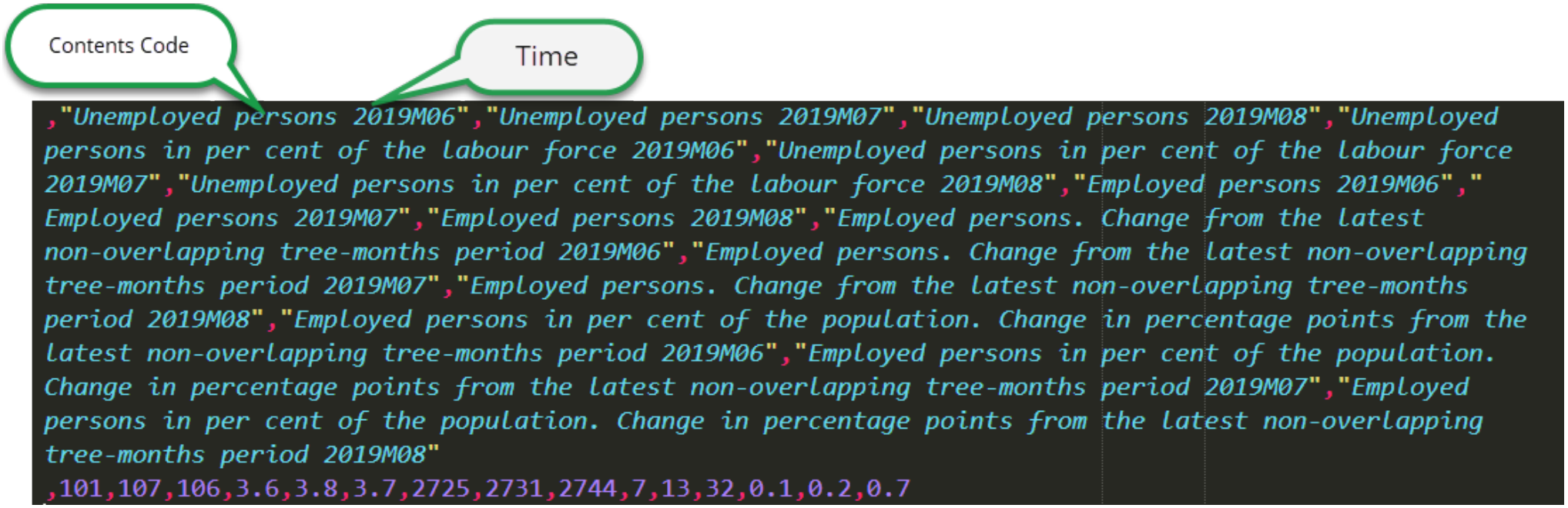
  - `"updated": "2019-06-24T06:00:00Z"`

- Same date as Excel output

533	Latest update:	
534	Gross operating expenditure for health and care services per inhabitant (NOK):	
535	20190624 08:00	

New local SN API output formats: CSV2 and CSV3

- What's wrong with today's CSV in PxWebApi?



- No pivot

Statistisk sentralbyrå
Statistics Norway

New
CSV2 - text

CSV3- codes

```
"month","contents","08930: Employment and unemployment, seasonally adjusted, 3-months moving average"
"2019M06","Unemployed persons",101
"2019M06","Unemployed persons in per cent of the labour force",3.6
"2019M06","Employed persons",2725
"2019M06","Employed persons. Change from the latest non-overlapping tree-months period",7
"2019M06","Employed persons in per cent of the population. Change in percentage points from the latest non-overlapping tree-months period",0.1
"2019M07","Unemployed persons",107
"2019M07","Unemployed persons in per cent of the labour force",3.8
"2019M07","Employed persons",2731
"2019M07","Employed persons. Change from the latest non-overlapping tree-months period",13
"2019M07","Employed persons in per cent of the population. Change in percentage points from the latest non-overlapping tree-months period",0.2
"2019M08","Unemployed persons",106
"2019M08","Unemployed persons in per cent of the labour force",3.7
"2019M08","Employed persons",2744
"2019M08","Employed persons. Change from the latest non-overlapping tree-months period",32
"2019M08","Employed persons in per cent of the population. Change in percentage points from the latest non-overlapping tree-months period",0.7
```

CSV2

```
"Tid","ContentsCode","08930"
"2019M06","Arbeidslause2",101
"2019M06","Arbeidslause4",3.6
"2019M06","Sysselsatt",2725
"2019M06","Sysselsette2",7
"2019M06","Sysselsette3",0.1
"2019M07","Arbeidslause2",107
"2019M07","Arbeidslause4",3.8
"2019M07","Sysselsatt",2731
"2019M07","Sysselsette2",13
"2019M07","Sysselsette3",0.2
"2019M08","Arbeidslause2",106
"2019M08","Arbeidslause4",3.7
"2019M08","Sysselsatt",2744
"2019M08","Sysselsette2",32
"2019M08","Sysselsette3",0.7
```

CSV3

API console

Table 1 08801: External trade in goods, by commodity number (HS) and country 1988 - 2017

Reset

Variables 2

- commodity number | Varekoder
- imports/exports | ImpEks
- country | Land
- contents | ContentsCode
- year | Tid

3 Show all metadata in new tab.

URL 4 https://data.ssb.no/api/v0/en/table/08801

Query 5

```
{
 "query": [
  {
   "code": "Varekoder",
   "selection": {
    "filter": "item",
    "values": [
     "00000011",
     "00000031",
     "00000032",
     "99319900",
     "99999999"
```

Open API query | Save API query 9 | 6 Run

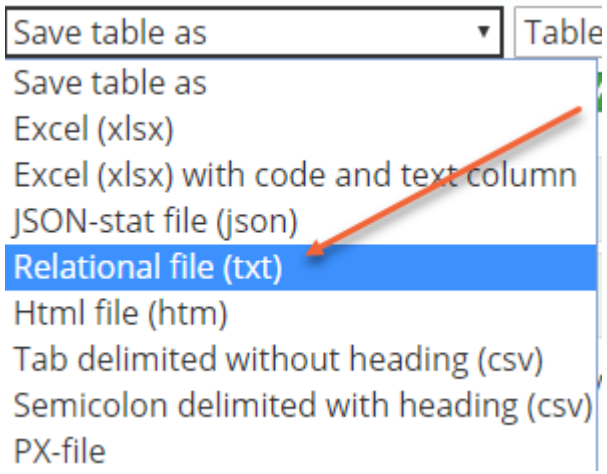
Status code 200 7

Result 8

```
    },
   "ImpEks": {
    "label": "imports/exports",
    "category": {
     "index": {
      "1": 0,
      "2": 1
     },
     "label": {
      "1": "Imports",
      "2": "Exports"
     }
    }
   },
   "Land": {
    "label": "country",
```

Statistisk sentralbyrå
Statistics Norway

Use of Relational file (txt)



- Pivot friendly format, data matrix - [Tidy data]

Tidy data

From Wikipedia, the free encyclopedia

Tidy data is an alternate name for the common statistical form called a *model matrix* or *data matrix*. A data matrix is defined in [1] as follows:

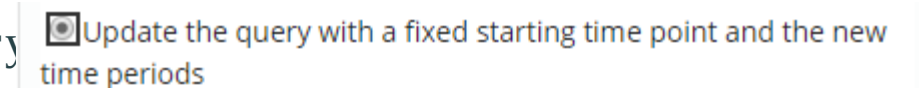
> A standard method of displaying a multivariate set of data is in the form of a data matrix in which rows correspond to sample individuals and columns to variables, so that the entry in the *i*th row and *j*th column gives the value of the *j*th variate as measured or observed on the *i*th individual.

Hadley Wickham later defined "Tidy Data" as data sets that are arranged such that each variable is a column and each observation (or *case*) is a row.[2] (originally with additional per-table conditions that made the definition equivalent to the Boyce–Codd 3rd normal form).

Data arrangement is an important consideration in data processing, but should not be confused with the also important task of data cleansing.

Other relevant formulations include denormalization prior to machine learning modeling (informally denoting moving data to a "wide form" where all possible measurements are in a given row), and use of semantic triples as intermediate representation (informally a "tall" or "long" form, where measurements about a single instance are spread across many rows).

Priorities for API development

1. Query improvements
   - Filter "*from*" – as in saved query 

   - mask single sign with "*?*" in filter "*all*"
   - Possibility to use filter all with * or *?* combined with *vs:* and *agg:*
2. JSON-stat2 - improvements
   - Footnotes
   - Extra date fields: retrieved, nest release (use extension)
3. New PxWebApi version 2

ERC – new service uses PxWebApi

- ERC - Economic Restful Client

  https://zhnzhn.github.io/

Hackathon at Statistics Norway

- Connection
  - combine and vizualize different tables in Python and R
- Hexystats
  - timeseries on maps in python
- Tutors
  - improvements i tutorials in Statbank and ssb.no
- Open sourcers
  - Experiences from contributing to a Open Source project
- Climate calculator
  - CO2 emission using different kinds of transportation
- TextMe
  - OCR on annual repoorts to get tables and footntes for business statistics, using R

TALL
SOM FORTELLER

Comments?

Statistisk sentralbyrå
Statistics Norway

Bonus tip:
Change MS Office language to English

Display and Help Languages

Within any Office application, select **File** > **Options** > **Language**.