

Mikael Nordberg
Statistikmyndigheten SCB

PxWebApi 1.0 beskrivning

Innehållsförteckning

PxWebApi 1.0 beskrivning	1
Innehållsförteckning.....	1
1 Läsanvisning	3
1.1 Vad denna instruktion omfattar	3
1.2 Hur detta dokument är uppbyggt.....	3
1.3 Terminologi.....	3
2 Komma igång	3
3 URL:er	3
3.1 API-NAME.....	4
3.2 API-VERSION.....	4
3.3 LANGUAGE	4
3.4 DATABASE-ID.....	4
3.5 LEVEL1..LEVELN.....	5
3.6 TABLE-ID	5
4 JSON format	5
4.1 Databaslistning	5
4.2 Nivåer för databaslistning.....	5
4.3 Metadatalistning för tabell	7
4.4 Uttagsfråga för tabell	8
4.4.1 Exempel på uttag i formatet CSV	10
4.4.2 Exempel på uttag i formatet CVS2	10

4.4.3	Exempel på uttag i formatet CSV3	10
4.5	Tabellsvar	11
4.5.1	Metadatadeln	11
4.5.2	Datadeln	11
5	Begränsningar	12
6	Loggning av användning	12

1 Läsanvisning

1.1 Vad denna instruktion omfattar

Denna instruktion beskriver användningen av PxWeb API:et. Instruktionen är främst riktad till personer som vill få en introduktion till hur API:et kan användas. Läsaren behöver inte ha någon erfarenhet av PxWeb för att förstå innehållet men det underlättar. Grundläggande kunskaper om http-protokollet underlättar också.

Observera att exemplen i denna instruktion är påhittade och inte nödvändigtvis fungerar i Statistikdatabasen på SCB.se.

1.2 Hur detta dokument är uppbyggt

Kapitel 3 och 4 riktar sig till slutanvändaren som vill lära sig mer om hur API:et kan användas. Övriga kapitel beskriver API:et ur ett driftperspektiv.

1.3 Terminologi

- HTTP verb – Kan ses som HTTP förfrågningsmetoder där GET och POST är de vanligaste.
- HTTP response code – HTTP-protokollet använder olika svarskoder för att indikera svarsstatusen från användarförfrågningarna. En lista med svarskoder finns här: <http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html>

2 Komma igång

Om du enbart vill bekanta dig med hur API:et används ska du läsa kapitel 3 och 4. Om du är intresserad av att exponera dina PC-Axis databaser bör samtliga kapitel läsas.

3 URL:er

Du når API:et via dess URL som på många vis påminner om ett RESTful API.

URL:en består av olika delar.

API-NAME/API-VERSION/LANGUAGE/DATABASE-ID/<LEVELS>/TABLE-ID

- API-NAME definierar rotnamnet för API:et.

- API-VERSION definierar versionen av API:et.
- LANGUAGE definierar språket för datahämtningen.
- DATABASE-ID definierar databasen där de statistiska tabellerna finns lagrade.
- LEVELS definierar noll eller flera nivåer av indelningar i databasen.
- TABLE-ID definierar identiteten för en statistisk tabell.

Exempel som visar URL:n tabellen BefProgFoddaMedel11:

```
/axis/v1/sv/ssd/BE/BE0401/BE0401B/BefProgFoddaMedel11
```

Föreslagen URL för Statistiska centralbyråns implementation av API:et skulle se ut:

```
http://api.scb.se/OV0104/v1/doris/en/ssd/BE/BE0401/BE0401B/BefProgFoddaMedel11
```

3.1 API-NAME

Om man ställer en fråga till API-NAME ska den antingen leda till en informationssida om API:et alternativt ett Http 404 Not Found-felmeddelande.

Statistiska centralbyråns API ger ett 404-felmeddelande.

3.2 API-VERSION

Om man ställer en fråga till /API-NAME/API-VERSION ska den antingen leda till en sida med information om användning för just den specifika versionen alternativt ett Http 404 Not Found-felmeddelande.

Statistiska centralbyråns API ger ett 404-felmeddelande.

3.3 LANGUAGE

Om man ställer en fråga till (HTTP verb GET) the /API-NAME/API-VERSION/LANGUAGE resulterar den i en JSON-formaterad svarssida som listar alla databaser för det angivna språket.

3.4 DATABASE-ID

Om man ställer en fråga till (HTTP verb GET) /API-NAME/API-VERSION/LANGUAGE/DATABASE-ID resulterar den i en JSON-formaterad svarssida som listar alla förstanivåer för den angivna databasen och språket.

Se 4.1 för resultatformatet.

3.5 LEVEL1..LEVELN

Om man ställer en fråga till (HTTP verb GET) /API-NAME/API-VERSION/LANGUAGE/DATABASE-ID/LEVEL1 till /API-NAME/API-VERSION/LANGUAGE/DATABASE-ID/LEVEL1../LEVELN resulterar den i en JSON-formaterad svarssida som listar alla undernivåer och/eller tabeller för den angivna nivån, databasen och språket.

Se 4.2 för resultatformatet.

3.6 TABLE-ID

Om man ställer en fråga till (HTTP verb GET) /API-NAME/API-VERSION/LANGUAGE/DATABASE-ID/LEVEL1../LEVELN/TABLE-ID resulterar den i en JSON-formaterad svarssida som listar metadata för den angivna tabellen.

Se 4.3 för resultatformatet.

Om man ställer en fråga till (HTTP verb POST) /API-NAME/API-VERSION/LANGUAGE/DATABASE-ID/LEVEL1../LEVELN/TABLE-ID krävs det att man skickar med en JSON-formaterad fråga som specificerar vilka data som ska hämtas för den aktuella tabellen.

Se 4.4 för den JSON-formaterade frågan och 4.5 för det JSON-formaterade svarsformatet.

4 JSON format

4.1 Databaslistning

Svaret består av en lista med databasobjekt som har en identitet, dbid och en textuell beskrivning, text.

Exempel resultat för “/doris/en” visar att vi har en databas.

```
[{"dbid":"ssd","text":"Statistics Sweden"}]
```

4.2 Nivåer för databaslistning

Svaret består av en lista med node-objekt som har en identitet, id, en textuell beskrivning, text, samt en typ, type. Det finns tre typer av node-objekt; l, t och h där l är en undernivå, t är en tabell och h är en rubrik.

Exempelresultatet för “/doris/en/ssd” visar att vi har 15 nivåer i den engelska databasen.

```
[
  {"id":"BE", "type":"l", "text":"Population"},
  {"id":"FM", "type":"l", "text":"Financial markets"},
  {"id":"HE", "type":"l", "text":"Household finances"},
  {"id":"LE", "type":"l", "text":"Living conditions"},
  {"id":"MI", "type":"l", "text":"Environment"},
  {"id":"NV", "type":"l", "text":"Business activities"},
  {"id":"PR", "type":"l", "text":"Prices and Consumption"},
  {"id":"AM", "type":"l", "text":"Labour market"},
  {"id":"BO", "type":"l", "text":"Housing, construction and
building"},
  {"id":"HA", "type":"l", "text":"Trade in goods and services"},
  {"id":"JO", "type":"l", "text":"Agriculture, forestry and
fishery"},
  {"id":"ME", "type":"l", "text":"Democracy"},
  {"id":"NR", "type":"l", "text":"National accounts"},
  {"id":"OE", "type":"l", "text":"Public finances"},
  {"id":"UF", "type":"l", "text":"Education and research"}
]
```

Exempelresultatet för “/doris/en/ssd/BE/BE0401/BE0401B” visar att vi har 4 tabeller för nivån BE0401B.

```
[
  { "id": "BefolkprognRev2009", "type": "t",
    "text": "Population by age and sex. Year 2009-2110 2009 -
2110"},
  { "id": "BefolkprognRev2010", "type": "t",
    "text": "Population by age and sex. Year 2010-2110 2010 -
2110"},
  { "id": "BefolkprognRev4", "type": "t",
    "text": "Population size by age and sex. Year 2005-2050
2004 - 2050"},
  { "id": "BefolkprognRev2008", "type": "t",
    "text": "Population by age and sex. Year 2008-2110 2008 -
2110"}
]
```

En rubriknod skulle kunna se ut så här:

```
[
  {"id":"XX", "type":"h", "text":"Population heading"}
]
```

4.3 Metadatalistning för tabell

Svaret består av en titel, title, samt en lista över variabler för tabellen. Variabelobjekten har fyra egenskaper; kod, namn, eliminering och tid (code, text, elimination, time).

Egenskapen för kod och namn är obligatoriska, medan de andra egenskaperna är frivilliga. Om inget värde är angivet för de frivilliga egenskaperna eliminering och tid antar dessa värdet falskt, *false*. Giltiga värden för dessa egenskaper är antingen sant eller falskt (true/false). Om en variabel har eliminering satt till sant måste man inte välja ett värde för den variabeln vid uttag.

Enbart en variabel kan ha egenskapen time satt till sant.

Variabelobjektet innehåller även två listor. En som listar alla koder för de värden som variabeln kan anta, values, samt en lista med de presentationstexter för de värden som variabeln kan anta.

Exempel resultat för URL:n

“/doris/sv/ssd/BE/BE0401/BE0401B/BefProgFoddaMedel10” visar att tabellen består av fyra variabler/dimensioner, Fodseland, Alder, ContentsCode och Tid.

```
{
  "title": "Births by country of birth, age and period",
  "variables":
  [
    {
      "code": "Fodseland",
      "text": "country of birth",
      "values": ["010", "020", "030", "040", "050", "060", "070"],
      "valueTexts":
      ["Sweden", "Nordic countries excl. Sweden",
      "EU excl. Nordic countries",
      "Europe excl. EU and Nordic countries", "Low HDI excl.
      Europe",
      "Medium HDI excl. Europe", "High HDI excl. Europe"],
      "elimination": true},
    {
      "code": "Alder",
      "text": "age", *
      "values": *
      ["-14", "15", "16", "17", "18", "19", "20", "21", "22", "23",
      "24", "25", "26", "27", "28", "29", "30", "31", "32", "33",
      "34", "35", "36", "37", "38", "39", "40", "41", "42", "43",
      "44", "45", "46", "47", "48", "49+"],
      "valueTexts":
      ["-14 years", "15 years", "16 years", "17 years",
      "18 years", "19 years", "20 years", "21 years",
      "22 years", "23 years", "24 years", "25 years",
      "26 years", "27 years", "28 years", "29 years",
      "30 years", "31 years", "32 years", "33 years",
      "34 years", "35 years", "36 years", "37 years",
      "38 years", "39 years", "40 years", "41 years",
      "42 years", "43 years", "44 years", "45 years",
      "46 years", "47 years", "48 years", "49+ years"],
      "elimination": true},
    {
      "code": "ContentsCode",
      "text": "observations",
      "values": ["BE0401M2"],
      "valueTexts": ["Births"]},
  ]
}
```

```

{
  "code": "Tid",
  "text": "period",
  "values":
  ["2010", "2011", "2012", "2013", "2014", "2015", "2016",
   "2017", "2018", "2019", "2020", "2021", "2022", "2023",
   "2024", "2025", "2026", "2027", "2028", "2029", "2030",
   "2031", "2032", "2033", "2034", "2035", "2036", "2037",
   "2038", "2039", "2040", "2041", "2042", "2043", "2044",
   "2045", "2046", "2047", "2048", "2049", "2050", "2051",
   "2052", "2053", "2054", "2055", "2056", "2057", "2058",
   "2059"],
  "valueTexts":
  ["2010", "2011", "2012", "2013", "2014", "2015", "2016",
   "2017", "2018", "2019", "2020", "2021", "2022", "2023",
   "2024", "2025", "2026", "2027", "2028", "2029", "2030",
   "2031", "2032", "2033", "2034", "2035", "2036", "2037",
   "2038", "2039", "2040", "2041", "2042", "2043", "2044",
   "2045", "2046", "2047", "2048", "2049", "2050", "2051",
   "2052", "2053", "2054", "2055", "2056", "2057", "2058",
   "2059"],
  "time": true
}
]
}

```

4.4 Uttagsfråga för tabell

Uttagsfrågan består av två delar, själva frågan samt en svarsdel.

Frågan består av en lista med selekteringsobjekt som specificerar vilka värden som är valda för varje variabel. Om ett selekteringsobjekt saknas för en variabel väljs dess värde automatiskt enligt följande regler:

1. Om variabeln går att eliminera (elimination=true) och har ett elimineringsvärde (en total) väljs enbart elimineringsvärdet.
2. Om variabeln går att eliminera (elimination=true) men saknar elimineringsvärde aggregeras samtliga värden för variabeln.
3. Om variabeln inte går att eliminera (elimination=false) väljs samtliga värden för variabeln.

Selektionsobjektet består av en egenskap för variabelkoden, code, och en selektering, selection. Selektionen har en filteregenskap som beskriver hur värden är angivna.

Filter som stöds är:

- **Item.** Detta filter listar alla värdekoder i en lista.
- **All.** Detta filter använder en så kallad wildcardselektering. Varje wildcardselektering ges i values listan. Enbart ett wildcard är tillåtet. T.ex. 01* ger alla värden som startar på 01, * ger alla värden.
- **Top.** Detta filter används för att välja de översta värdena. Antalet värden som ska väljas ges som ett värde i values listan. Det får enbart förekomma ett värde i values listan och det måste vara ett positivt tal större än 0. Om variabeln är en tidsvariabel väljs de senaste tidspunkterna.

- Agg. Detta filter visar att värdena i values listan är aggregerade värden. Identiteten för aggregeringen anges efter ett kolon i filtret t.ex. agg:age5.
- Vs. Detta filter visar att värdena i values listan är från en annan värdemängd. Identiteten för värdemängden anges efter ett kolon i filtret t.ex. vs:NUTS.

Svarsobjektet, response, är frivilligt att ange. Om det inte specificeras kommer resultatet som en PC-Axis PX fil.

Format som stöds är

- px
- csv
- csv2
- csv3
- json
- xlsx
- json-stat
- json-stat2
- sdmx

Formatet json-stat returnerar svaret som JSON-stat version 1.2 och formatet json-stat2 returnerar svaret som JSON-stat version 2.0.

Exempel POST-frågan nedan till URL:n

“/doris/en/ssd/BE/BE0401/BE0401B/BefProgFoddaMedel10”. Frågan specificerar alla värden för varje variabel som vi är intresserade av.

```
{
  "query": [{"code": "Fodelseland",
    "selection": {"filter": "item",
    "values": ["010", "020"]}},
    {"code": "Alder", "selection": {"filter": "all",
    "values": ["*"]}},
    {"code": "Tid", "selection": {"filter": "top",
    "values": ["3"]}}],
  "response": {"format": "csv"}
}
```

4.4.1 Exempel på uttag i formatet CSV

Inledande rad med variabelnamn och innehåll. Klartexter på följande rader.

```
"region","civilstånd","ålder","kön","Folkmängd 2023"  
"01 Stockholms län","ogifta","20-29 år","män",142540  
"01 Stockholms län","ogifta","20-29 år","kvinnor",127966  
"01 Stockholms län","ogifta","30-39 år","män",127966  
"01 Stockholms län","ogifta","30-39 år","kvinnor",104126  
"01 Stockholms län","ogifta","40-49 år","män",63877  
"01 Stockholms län","ogifta","40-49 år","kvinnor",50624  
"01 Stockholms län","ogifta","50-59 år","män",45647  
"01 Stockholms län","ogifta","50-59 år","kvinnor",39319
```

4.4.2 Exempel på uttag i formatet CVS2

Tabellen är pivoterad så att alla variabler visas i den inledande raden. Följande rader visar klartext för variablerna och ett värde per rad.

```
"region","civilstånd","ålder","kön","år","tabellinnehåll","Folkmängd"  
"01 Stockholms län","ogifta","20-29 år","män","2023","Folkmängd",142540  
"01 Stockholms län","ogifta","20-29 år","kvinnor","2023","Folkmängd",127966  
"01 Stockholms län","ogifta","30-39 år","män","2023","Folkmängd",127966  
"01 Stockholms län","ogifta","30-39 år","kvinnor","2023","Folkmängd",104126  
"01 Stockholms län","ogifta","40-49 år","män","2023","Folkmängd",63877  
"01 Stockholms län","ogifta","40-49 år","kvinnor","2023","Folkmängd",50624  
"01 Stockholms län","ogifta","50-59 år","män","2023","Folkmängd",45647  
"01 Stockholms län","ogifta","50-59 år","kvinnor","2023","Folkmängd",39319
```

4.4.3 Exempel på uttag i formatet CSV3

Den inledande raden innehåller variabelnamn samt tabellens unika id. Följande rader visar koder för variablerna och ett värde per rad.

```
"Region","Civilstand","Alder","Kon","Tid","ContentsCode","TAB638"  
"01","OG","20-29","1","2023","BE0101N1",142540  
"01","OG","20-29","2","2023","BE0101N1",127966  
"01","OG","30-39","1","2023","BE0101N1",127966  
"01","OG","30-39","2","2023","BE0101N1",104126  
"01","OG","40-49","1","2023","BE0101N1",63877  
"01","OG","40-49","2","2023","BE0101N1",50624  
"01","OG","50-59","1","2023","BE0101N1",45647  
"01","OG","50-59","2","2023","BE0101N1",39319
```

4.5 Tabellsvår

Tabellsvår är indelat i en metadatadel och en datadel. Metadatadeln består av en lista med kolumner som listar olika dimensioner och innehåll för en tabell, liksom kommentarer som kan finnas för en datatabell. Datadeln består av de mätvärden som finns för tabellen.

4.5.1 Metadatadeln

Kolumnobjektet består av 5 egenskaper:

- code. Identitet för kolumnen.
- text. En textuell beskrivning av kolumnen
- type. Typ av kolumn. Den har tre möjliga värden:
 - *d* – en dimension. Detta är defaultvärdet om det inte skulle vara specificerat
 - *t* – en tidsdimension
 - *c* – en innehållskolumn
- unit. Om typen är “c” kan en enhet specificeras som beskriver enheten för värdet som anges i datadeln. Denna egenskap ska ignoreras om inte type=c
- comment. En frivillig kommentar för kolumnen/dimensionen.

4.5.2 Datadeln

Datadeln ges i form av nyckel och värden. Nyckeln specificerar dimensionerna och värden specificerar mätvärdena för dimensionerna. Ordningen som värden i nyckeln anges på ska vara samma ordning som de anges i kolumnlistan. Likaså är ordningen i values-listan densamma som kolumnerna av type=c i kolumnlistan.

Det finns även en frivillig kommentarslista för varje nyckel-värdepar som specificerar en kommentar för ett värde. Indexet i listan motsvarar indexet för mätvärdet i values-listan.

```
{
  "columns": [{ "code": "region", "text": "Region" },
    { "code": "ageG5", "text": "Age",
      "comment": "Citizens under the age of 6 are not..." },
    { "code": "period", "text": "Time", "type": "t" },
    { "code": "x", "text": "Population", "type": "c",
      "unit": "amount" } ],
  "comments": [ { "variable": "period", "value": "2005",
    "comment": "Preliminary figures" } ],
  "data": [ { "key": [ "02", "0-7", "2003" ], "values": [ 100 ] },
    { "key": [ "02", "0-7", "2004" ], "values": [ 101 ] },
    { "key": [ "02", "0-7", "2005" ], "values": [ 102 ] },
    { "key": [ "02", "8-14", "2003" ], "values": [ 103 ] },
    { "key": [ "02", "8-14", "2004" ], "values": [ 104 ] },
    { "key": [ "02", "8-14", "2005" ], "values": [ 105 ] },
    { "key": [ "02", "15-23", "2003" ], "values": [ 106 ] },
    { "key": [ "02", "15-23", "2004" ], "values": [ 107 ] },
    { "key": [ "02", "15-23", "2005" ], "values": [ 108 ] },
    { "key": [ "02", "24-33", "2003" ], "values": [ 109 ] },
    { "key": [ "02", "24-33", "2004" ], "values": [ 110 ] },
    { "key": [ "02", "24-33", "2005" ], "values": [ 111 ] },
    { "key": [ "02", "34-54", "2003" ], "values": [ 112 ] } ]
}
```

```

{"key": ["02", "34-54", "2004"], "values": [113]},
{"key": ["02", "34-54", "2005"], "values": [114]},
{"key": ["02", "55+", "2003"], "values": [115]},
{"key": ["02", "55+", "2004"], "values": [116]},
{"key": ["02", "55+", "2005"], "values": [117]},
{"key": ["06", "0-7", "2003"], "values": [118]},
{"key": ["06", "0-7", "2004"], "values": [119]},
{"key": ["06", "0-7", "2005"], "values": [120]},
{"key": ["06", "8-14", "2003"], "values": [121]},
{"key": ["06", "8-14", "2004"], "values": [122]},
{"key": ["06", "8-14", "2005"], "values": [123]},
{"key": ["06", "15-23", "2003"], "values": [124]},
{"key": ["06", "15-23", "2004"], "values": [125]},
{"key": ["06", "15-23", "2005"], "values": [126]},
{"key": ["06", "24-33", "2003"], "values": [127]},
{"key": ["06", "24-33", "2004"], "values": [128]},
{"key": ["06", "24-33", "2005"], "values": [129]},
{"key": ["06", "34-54", "2003"], "values": [130]},
{"key": ["06", "34-54", "2004"], "values": [131]},
{"key": ["06", "34-54", "2005"], "values": [132]},
{"key": ["06", "55+", "2003"], "values": [133]},
{"key": ["06", "55+", "2004"], "values": [134]},
"comment": ["Imputed"]},
{"key": ["06", "55+", "2005"], "values": [135]},
}

```

5 Begränsningar

Användandet av API kan begränsas. Begränsningsfunktionen är baserat på anroparens IP-adress. Det fungerar på så sätt att anroparen kan göra 10 anrop på ett tidsfönster på 10 sekunder (dessa värden kan komma att konfigureras i efterhand).

Tidsperioden 10 sekunder är en glidande tidsperiod.

Om anroparen gör fler anrop än tillåtet skickas ett **HTTP 429 Too Many Requests** svar.

6 Loggning av användning

Loggramverket log4net används för loggning av alla frågor som ställs mot API:et. API:et använder **api-usage** loggern för att logga användningen. Information som loggas är anroparens IP-adress, tabellen som hämtas samt antalet mätvärden som hämtas.