

CENTRĀLĀ STATISTIKAS PĀRVALDE

Datu publicēšana Oficiālās statistikas portālā

Informācija iestādēm, kuru dati iekļauti oficiālās statistikas programmā

Sadarbība datu sagatavošanā un publicēšanā

Portālā publicētā statistika

1

Atbilst atvērto datu principiem

- Lejupielādējami atvērtā, mašīnlasāmā formātā (CSV, JSON u.c.)
- Pieejami arī caur API un MI pakalpes
- Izmanto vienotus identifikatorus (URI)
- Bez maksas, bez reģistrēšanās

2

Augsts analītiskais potenciāls

- Iespējami garas laikrindas
- Detalizētas, daudzdimensiju tabulas
- Rādītāji standartizēti kodēti visu iestāžu datus
- Automātiski atjaunotas diagrammas un kartes

3

Viegli atrodams

- Indeksēts Google meklētājā
- Portāla saturs pieejams plašai auditorijai

4

Divvalodu pieejamība

- Informācija tiek publicēta arī angļu valodā

Oficiālās statistikas portāls

- **Vienota piekļuves vieta** visai publicētajai oficiālajai statistikai (stat.gov.lv)
- **Vienota vide** CSP un citu oficiālās statistikas iestāžu datu publicēšanai
- **Balstīta uz PxWeb** – starptautiski izplatītu daudzdimensionālu statistikas tabulu sistēmu
- **Vienots metodoloģijas standarts** un datu tabulu struktūra visām iestādēm



Standartizēts un mašīnlasāms saturs

Katra tabula ir vienlaikus cilvēklasāma un mašīnlasāma – ar vienotu struktūru un atvērtu API

Katra tabula ir API galapunkts

Datubāze pieejama caur PxWeb API v1 vai v2. Datus var pieprasīt XLSX, CSV, JSON-stat2 un PX formātā.

```
# Pieprasījuma piemērs ar API v1  
https://data.stat.gov.lv/api/v1/lv/OSP\_PUB/START/EMP/NBBA/NVA/NVA011m
```

```
# Pieprasījuma piemērs ar API v2  
https://api.stat.gov.lv/api/v2/tables/NVA011m/data?lang=lv&valueCodes\[ContentsCode\]=NVA011m1,NVA011m&valueCodes\[TIME\]=2026M04&outputformat=json-stat2
```

Strādā ar CSP datiem tieši R un Python vidē

Tā kā PxWeb API tiek plaši izmantots, ir pieejamas gatavas bibliotēkas

R

R ekosistēma

pxweb

Tiešs PxWeb API klients (pakotne ļauj pieprasīt CSP datus tieši no koda)

```
# 1. Interaktīvs vaicājuma veidotājs
library(pxweb)
d <- pxweb_interactive("https://data.stat.gov.lv/api/v1/lv/")

# 2. Vai tieši ar vaicājumu
pxq <- pxweb_query(list("AREA" = c("LV0023000"), "TIME" = c("2024")))
pxd <- pxweb_get(url = ".../IRS031", query = pxq)

# 3. Pārvērš par data.frame
df <- as.data.frame(pxd, column.name.type = "text",
variable.value.type = "text")
# → df: novads | gads | vērtība
```

Py

Python ekosistēma

pxstat

PxWeb API klients – analogs R pxweb

pandas

Datu rāmji un analīze

requests

Tiešs API izsaukums

```
# 1. Ar pxstat klientu (ieteicams)
from pxstat import pxweb
px = pxweb.Pxweb("https://data.stat.gov.lv/api/v1/lv/")
df = px.get("OSP_PUB/.../IRS031", {"AREA": ["LV0023000"], "TIME":
["2024"]})
# → DataFrame: novads | gads | vērtība

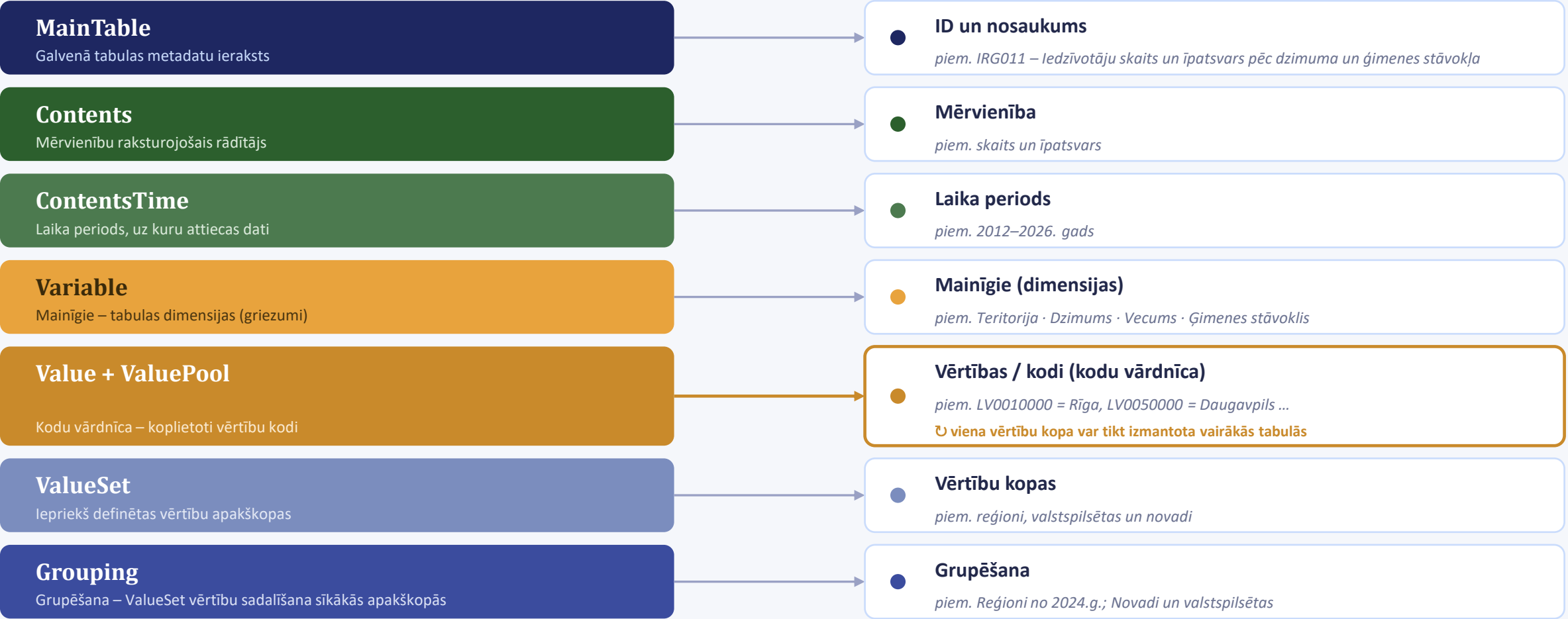
# 2. Tieši ar requests
import requests, pandas as pd
r = requests.post(url, json=q)
df = pd.DataFrame(r.json())
```

Kā tabulas struktūra top no metadatiem

Katrs publiskotās tabulas elements nāk no citas metadatu tabulas – noderīgi zināt, gatavojot datus iesniegšanai

METADATU TABULA

PUBLISKOTAJĀ TABULĀ KĻŪST PAR



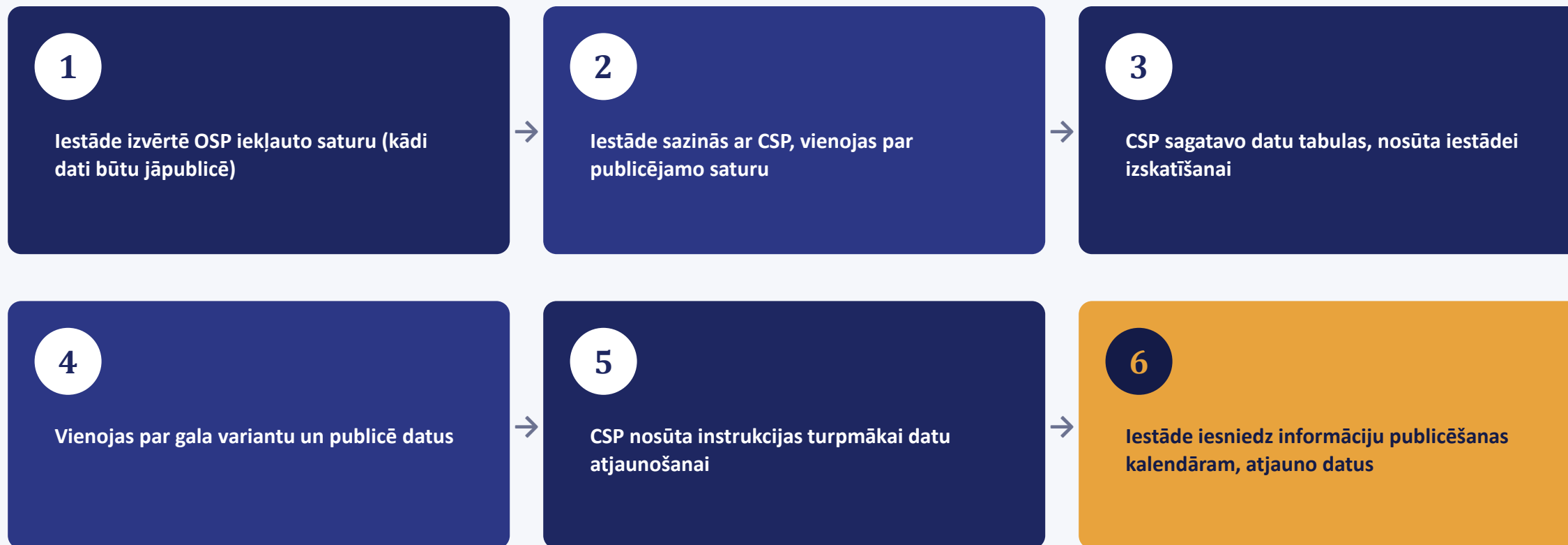
Vienoti kodi un kodu vārdnīcas

Bez vienotiem kodiem katrs datu iesniedzējs varētu lietot atšķirīgus apzīmējumus vienām un tām pašām parādībām – **kodu vārdnīca ir centralizēts references punkts, uz kuru atsaucas visas tabulas.**



Rādītāji – standartizētas vērtības, kas sastāv no koda un rādītāja nosaukuma 2 valodās, apkopotas pēc klasifikācijas piederības atsevišķās kodu vārdnīcās. Atkārtoti tiek izmantotas dažādās tabulās visā datubāzē. Kodu vārdnīcas ir kopīga valoda, ko izmanto visas portālā publicētās tabulas.

Veicamie soļi jaunu datu publicēšanai portālā



Datu ielādes kanāli

Manuālā ielāde

Izmantojot portāla saskarni

- Izpilde caur administrēšanas saskarni
admin.stat.gov.lv
- Var ielādēt atsevišķus failus: CSV un Excel

Automatizēta ielāde / API

Pielāgots risinājums

- Iespējami arī API risinājumi datu iesniegšanai
- Konkrētais risinājums tiek pieskaņots iestādes tehniskajām iespējām un vajadzībām

Neatkarīgi no izvēlētā kanāla, dati jāimportē iepriekš definētā struktūrā un jāpārbauda pirms publicēšanas.

Validācija

Validācijas process palīdz identificēt kļūdas un neatbilstības, novēršot nekorektu datu publicēšanu un norādot datu piegādātājam uz vietām, kurās nepieciešami labojumi.

Datu ielādes procesā, sistēma pārbauda datu failā esošās informācijas atbilstību sistēmā definētajam:

- kolonnu nosaukumi un struktūra
- rādītāju kodi
- laika vienība (periodiskums, pieraksts)
- dati (pieļaujamās vērtības)
- nosacītie apzīmējumi (atbilstība sistēmā definētajiem)
- atbilstība kalendāram (vai tiek importēts periods, kas norādīts kalendārā)
- vērtības, kas atkārtojas

Biežāk sastopamās kļūdas datu importēšanas procesā

Tabulā izmantoti rādītāji, kas nav pievienoti sistēmā

- Failā ir norādīts rādītāja kods, kas nav pievienots tabulai piesaistītajā kodu apakšvārdnīcā
- Rādītāja kods ir atšķirīgs no sistēmā definētā

"Tukšās rindas" Excel failā

- Excel failā ir bijis saturs, kas vēlāk ir izdzēsts no datu šūnām

Nepareizs decimāldaļu atdalītājs Excel failā

- Decimāldaļu atdalītājam jābūt punktam

Neatbilstošs periods vai periods neatbilst kalendārā definētajam

- Importējamais periods neatbilst sistēmā definētajam periodiskumam
- Kalendārā definēts cits periods kā importa failā

Dublēti ieraksti – vienam un tam pašam rādītājam un periodam failā ir vairāk nekā viena vērtība.

Svarīgi atcerēties



Saziņai izmantojam **portals@csp.gov.lv**



Pēc datu publicēšanas, **iesniedzam nākamo plānoto datu publicēšanas datumu**



Sekojam līdz **publicēšanas datumiem, laicīgi ievieojam datus sistēmā**

Kopsavilkums

Veiksmīga datu publicēšana portālā balstās uz labu sadarbību starp CSP un datu iesniedzēju iestādi:

Vienoti standarti

kodi un struktūra visai datubāzei

Skaidrs process

solis pa solim, no sagatavošanas līdz
publicēšanai

Rūpīga validācija

kvalitātes pārbaude pirms
publicēšanas

Jautājumi? portals@csp.gov.lv

MCP: <https://mcp-px.stat.gov.lv/mcp>