



Datum
23-03-2021
Versie 1.2

Van
M. Dekker

Aan:
Gebruikers open match data UWV WERKbedrijf

Ons kenmerk
UWVopenmatchdata

Uw kenmerk
-

Onderwerp
Specificatie

Kopie aan
-

Bijlage(n)
-

Specificatie 'open match data' UWV WERKbedrijf

Attentie bij databestand 'UWV Open Match Data 16-03-21':

Het databestand van week 11-2021 bevat een afwijkend, fors hoger, aantal vacatures ten opzichte van de weken daar voor. Een andere verklaring dan dat er sprake is van een incidentele piek in de aanlevering van vacatures is er niet. Bij ongewenste uitkomsten dan svp deze ene week buiten beschouwing laten.

Wijziging per 17-12-2019:

De bestanden die per 17-12-2019 gepubliceerd worden zijn op het volgende aspect verbeterd:

- In de oude levering, gepubliceerd voor 10-12-19, waren cv's waar de vervaldatum was verstreken niet meegenomen. Dat blijkt niet conform het uitvoeringsproces te zijn. Deze cv's staan nog wel op actief en worden in de match meegenomen. Deze omissie is nu hersteld.
- De draaidatum is een dag opgeschoven.

De bestanden gepubliceerd voor 10-12-19 zijn daardoor in feite tenminste onvolledig en zijn daarom niet goed bruikbaar. Het betreft de bestanden met bestandsnamen:

- UWV Open Match Data 25-11-19 ZIP
- UWV Open Match Data 02-12-19 ZIP
- UWV Open Match Data 09-12-19 ZIP

De correcte bestanden, gepubliceerd op/na 17-12-19, zijn:

- UWV Open Match Data 26-11-19 ZIP
 - UWV Open Match Data 03-12-19 ZIP
 - UWV Open Match Data 10-12-19 ZIP
 - UWV Open Match Data 17-12-19 ZIP
 - Etc etc
-

1. Inleiding

Dit document beschrijft de gegevensset die door UWV WERKbedrijf als open data beschikbaar wordt gesteld voor gebruik door eenieder die daarvoor belangstelling heeft. Deze data zijn gebaseerd op de (geanonimiseerde) beroepen informatie uit cv's en vacatures van Werk.nl.

De 'open match data' set bevat voor elk van de ruim 3.200 actuele beroepen, de zogenaamde referentieberoepen, die UWV WERKbedrijf onderkent informatie over de arbeidsmarkt voor een opgegeven locatie.

Het vrijgeven van de data maakt het mogelijk om deze naar eigen inzicht te verwerken in andersoortige producten, al dan niet in combinatie met gegevens uit andere bronnen, bijvoorbeeld het CBS.

2. Informatie en bewerkingen voor bruikbare open data

De gegevens onder de 'open match data' set is samengesteld op basis van vacatures en *geanonimiseerde* cv's op werk.nl. Voor de 'open match data' worden deze geaggregeerd 1) per beroep en 2) per viercijferig postcodegebied. Voor beroepen kan gelden dat er per postcodegebied minder werkzoekenden zijn dan de drempel van 10 personen die UWV heeft vastgesteld. Voor de 'open match data' set levert dit veelal geen problemen op omdat er nagenoeg altijd wordt gezocht in een ruimer gebied. Om te voorkomen dat alleen de meest populaire beroepen in de dataset terechtkomen worden de gegevens niet alleen per postcodegebied verstrekt maar ook per gemeente, per provincie en landelijk. Hierdoor komt toch veel informatie beschikbaar die op het niveau van een postcodegebied onder de drempel blijft.

Per beroep en geografisch gebied worden drie soorten informatie verstrekt:

1. **Informatie over gewenste beroepen (W):** Elke werkzoekende in werk.nl heeft 1 tot 3 beroepen opgegeven waarin hij of zij specifiek is geïnteresseerd. Als er meerdere beroepen zijn die tot hetzelfde referentieberoep horen, worden deze als '1' geteld.
2. **Informatie over beroepen waarin de werkzoekende ervaring heeft opgedaan (E):** Elke werkzoekende in werk.nl kan zijn/haar volledige CV opvoeren.
3. **Vacature informatie (V):** Hiervoor geldt geen drempel en deze worden daarom alleen per postcodegebied gegeven. Aggregaties naar gemeentelijk, provinciaal of landelijk niveau kunnen zelf worden gemaakt.

De verstrekte informatie geldt steeds voor één moment in de tijd. UWV voert zelf een registratie op weekbasis.

De beroepen waarover wordt gerapporteerd zijn zogenaamde *referentieberoepen*. De meeste van deze ca. 3.200 beroepen kennen nog een hoger aggregatieniveau, de zogenaamde *beroepscluster*. Bij de referentieberoepen wordt het daarbij horende beroepscluster geselecteerd en deze komt als apart veld in de rapportage terug; er zijn echter een, beperkt, aantal referentieberoepen die géén beroepscluster hebben en in dat geval zal het veld beroepscluster dan ook leeg blijven. Nadere informatie over de referentieberoepen is te vinden op <https://register.boc-uwv.nl>. Anders dan het geval is bij de geografische indeling wordt er hier geen gebruik gemaakt van het tonen van data die op het niveau van een individueel registerberoep onder de drempel van 10 personen vallen.

Zowel de statistieken als de bij behorende referentiegegevens zijn opgenomen in hetzelfde bestand. Dit bestand bevat dus redundantie met betrekking tot:

1. Beroepen: code en omschrijving van zowel het referentieberoep als het beroepscluster zijn opgenomen in het datarecord. Bij eenzelfde code referentieberoep komen steeds dezelfde beroepsnamen en dezelfde beroepscluster code voor.
2. Geografie: nummer postcodegebied, gemeente- en provincienaam zijn opgenomen in het datarecord. Bij dezelfde postcode wordt steeds dezelfde gemeente gevonden en bij eenzelfde gemeente steeds dezelfde provincie. Ontbreken al deze gegevens dan betreft het een landelijke telling.

Het voordeel van het combineren van data en referentiegegevens in hetzelfde bestand is tweemaal:

1. Minder technische gebruikers hebben de referentiegegevens bij de hand terwijl ICT experts de referentiegegevens gemakkelijk kunnen afleiden uit de data.
2. Veranderingen in de beroepentabel en in de geografische indeling van Nederland zijn steeds *up to date*.

Als nadelen kunnen worden genoemd dat het gecombineerde bestand groter is en dat referentiegegevens alleen worden geleverd wanneer ze relevant zijn voor de geleverde dataset. Zo zijn beroepen waarvoor geen vacatures bestaan en die ook landelijk onder de 10 werkzoekenden blijven niet in de gegevensset terug te vinden.

3. De pakbon (UWVopenmatchdatapakbonYYYYMMDD.txt)

Elke levering bevat een pakbonbestand zodat globaal duidelijk is welke data beschikbaar zijn. Voor de vrijgave van de gegevens op 18-11-2019 bevat het pakbonbestand (UWVopenmatchdatapakbon20191125.xls) de volgende informatie over het databestand (UWVopenmatchdata20191125.csv):

peildatum	categorie	groepering	records	beroepen	rectypen	geotypen
25-11-2019	01-Totalen	Peildatum	146337	2524	3	4
25-11-2019	02-Recordtype	ErvaringsBeroep	22832	2020	1	4
25-11-2019	02-Recordtype	Vacature	112977	2219	1	1
25-11-2019	02-Recordtype	WensBeroep	10528	1479	1	4
25-11-2019	03-Geotype	Gemeente	14384	632	2	1
25-11-2019	03-Geotype	Landelijk	3499	2037	2	1
25-11-2019	03-Geotype	Postcode	116885	2219	3	1
25-11-2019	03-Geotype	Provincie	11569	1297	2	1
25-11-2019	04-Rec-/Geotype	Ervaringsberoep-Gemeente	10115	624	1	1
25-11-2019	04-Rec-/Geotype	Ervaringsberoep-Landelijk	2020	2020	1	1
25-11-2019	04-Rec-/Geotype	Ervaringsberoep-Postcode	3265	65	1	1
25-11-2019	04-Rec-/Geotype	Ervaringsberoep-Provincie	7432	1278	1	1
25-11-2019	04-Rec-/Geotype	Vacature-Postcode	112977	2219	1	1
25-11-2019	04-Rec-/Geotype	Wensberoep-Gemeente	4269	342	1	1
25-11-2019	04-Rec-/Geotype	Wensberoep-Landelijk	1479	1479	1	1
25-11-2019	04-Rec-/Geotype	Wensberoep-Postcode	643	14	1	1
25-11-2019	04-Rec-/Geotype	Wensberoep-Provincie	4137	794	1	1

De pakbon bevat informatie over de peildatum, tellingen van het aantal records, het aantal verschillende beroepen, het aantal recordtypen (maximaal 3: ervaringsberoepen, wensberoepen en vacatures) en het aantal geografische ingangen (maximaal 4: postcodegebieden, gemeenten, provincies, landelijk).

De verdere dataregels op de pakbon geven dezelfde informatie uitgesplitst naar recordtypen, geografie en de combinatie van recordtype en geografie.

Direct wordt duidelijk dat de data op postcodeniveau voor wat betreft de cv's slechts voor een handvol beroepen bruikbaar is vanwege de gehanteerde drempel.

4. Het 'open match data' bestand (UWVopenmatchdataYYYYMMDD.csv)

Dit bestand heeft de volgende structuur:

#	Rubrieknaam	Betekenis
1	PEILDATUM	Peildatum gegevens (voor alle records identiek)
2	BEROEP_CD	Code referentieberoep UWV
3	REFERENTIEBEROEP	Beroepnaam referentieberoep UWV
4	BEROEPENCLUSTER_CD	Code Beroepscluster UWV
5	BEROEPENCLUSTER	Beroepnaam Beroepscluster UWV
6	POSTCODEGEBIED	Numeriek deel postcode
7	PROVINCIE	Provincienaam
8	GEMEENTE	Gemeentenaam
9	REC_TYPE	Record type (Ervaringsberoep, Wensberoep, Vacature)
10	AANTAL	Gecumuleerd aantal wens- en ervaringsberoepen uit cv en beroepen uit Vacature
11	GEM_LEEFTIJD	Gemiddelde leeftijd
12	AANTAL_MAN	Aantal mannen
13	AANTAL_VROUW	Aantal vrouwen
14	OPLNIV_GEM	Gemiddeld opleidingsniveau
15	ZOEKSTRAAL_GEM	Gemiddelde zoekstraal in kilometers
16	AANT_JR_ERV_GEM	Gemiddeld aantal jaar ervaring
17	AANT_LEEFT_TOT_27	Aantal met leeftijd t/m 27 jaar
18	AANT_LEEFT_28_39	Aantal met leeftijd 28 t/m 39 jaar
19	AANT_LEEFT_40_49	Aantal met leeftijd 40 t/m 49 jaar
20	AANT_LEEFT_50_PLUS	Aantal met leeftijd 50 jaar en ouder
21	AANT_LEEFT_ONBEKEND	Aantal met onbekende leeftijd
22	AANT_UPW_0_8	Aantal met maximaal 8 uur per week
23	AANT_UPW_9_16	Aantal met 9 t/m 16 uur per week
24	AANT_UPW_17_24	Aantal met 17 t/m 24 uur per week
25	AANT_UPW_25_32	Aantal met 25 t/m 32 uur per week
26	AANT_UPW_33_PLUS	Aantal met 33 of meer uur per week
27	AANT_UPW_ONBEKEND	Aantal met onbekend aantal uur per week
28	AANT_OPLNIV_1	Aantal met opleidingsniveau 1
29	AANT_OPLNIV_2	Aantal met opleidingsniveau 2
30	AANT_OPLNIV_3	Aantal met opleidingsniveau 3
31	AANT_OPLNIV_4	Aantal met opleidingsniveau 4
32	AANT_OPLNIV_5	Aantal met opleidingsniveau 5
33	AANT_OPLNIV_6	Aantal met opleidingsniveau 6
34	AANT_OPLNIV_ONBEKEND	Aantal met opleidingsniveau onbekend
35	AANT_WE_0_1	Aantal met minder dan 1 jaar werkervaring
36	AANT_WE_1_3	Aantal met 1 tot 3 jaar werkervaring
37	AANT_WE_3_6	Aantal met 3 tot 6 jaar werkervaring
38	AANT_WE_7	Aantal met 7 of meer jaar werkervaring
39	AANT_WE_ONBEKEND	Aantal met werkervaring onbekend

#	Rubrieknaam	Betekenis
---	-------------	-----------

De combinatie van de rubrieken 'beroep_cd', 'rec_cd' en 'postcode/gemeente/provincie' is uniek. Worden bestanden samengevoegd voor longitudinale analyse dan dient de rubriek peildatum te worden toegevoegd.

De betekenis van de gegevenselementen spreekt voor zich.

Uitzondering zijn de opleidingsniveaus waaraan de volgende betekenis wordt toegekend:

Code	Betekenis
------	-----------

- | | |
|---|----------------|
| 0 | Niet bekend |
| 1 | Basisonderwijs |
| 2 | VMBO |
| 3 | Havo/VWO |
| 4 | MBO-algemeen |
| 5 | HBO/bachelor |
| 6 | WO/master |

Een gemiddeld opleidingsniveau van 4,32 duidt dus aan dat het tussen MBO en HBO ligt maar dichterbij MBO.