



Datum
14 januari 2018
Versie 1.2

Van
M. Dekker

Aan: Gebruikers Open Data UWV Beroepenkaart

Ons kenmerk
UWVopenbkaartdata

Uw kenmerk
-

Onderwerp
Specificatie

Kopie aan
-

Bijlage(n)
-

Specificaties open data UWV WERKbedrijf beroepenkaart

Versie: 1.3, 29 mei 2019

Vanwege de grote vraag naar actuelere data in deze dataset wil UWV dit gaan bewerkstellingen. Op dit moment is de techniek om structureel te actualiseren niet robuust genoeg. Alles is in het werk gesteld om de techniek te optimaliseren zodat deze voldoende leveringsbetrouwbaar is. Zodra dit is gelukt zal er geactualiseerd worden.

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 beroepenkaart van Werk.nl (<https://spiegel.werk.nl/werkgever/beroepenkaart>). De beroepenkaart bevat voor elk van de ruim 3.200 actuele beroepen 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. Ook kunnen eenvoudig trends worden afgeleid, wat met de beroepenkaart zelf niet kan.

2. Informatie en bewerkingen voor bruikbare open data

De gegevens onder de beroepenkaart zijn vacatures en *geanonimiseerde* CVs in werk.nl. Voor de beroepenkaart 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 beroepenkaart 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.
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 onderliggende database (SPIEGEL) is historisch zodat de onderliggende detailgegevens blijvend bij UWV beschikbaar zijn.

De beroepen waarover wordt gerapporteerd zijn zogenaamde *registerberoepen*. De meeste van deze ca. 3.200 beroepen kennen nog een hoger aggregatieniveau, de zogenaamde *koepelberoepen*. Deze koepelberoepen bieden de mogelijkheid om op een hoger niveau te aggregeren. Nadere informatie over de registerberoepen 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 redundantie¹ met betrekking tot:

1. **Beroepen:** code en omschrijving van zowel het registerberoep als het koepelberoep zijn opgenomen in het datarecord. Bij eenzelfde code registerberoep komen steeds dezelfde beroepsnamen en dezelfde koepelberoep 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 tweërlei:

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.

Let op: het komt voor dat de aantallen 'aantal uren per week' hoger zijn dan de aantallen mensen met ervaring met het bijbehorende beroep. Het probleem blijkt voort te komen uit de

¹ Voor technisch georiënteerde lezers: het bestand is in 1^e normaalvorm.

omstandigheid dat mensen hetzelfde beroep meermalen hebben uitgeoefend. Bijvoorbeeld als iemand uit Assen het beroep van groepsleerkracht basisonderwijs meerdere malen heeft uitgeoefend dan worden diens werkervaringen geteld en komt de totaalstelling van de 'aantal uren per week' rubrieken dus (ruim) boven het aantal unieke CV's uit.

3. De pakbon (UWVopenbkaartpakbonYYYYMMDD.txt)

Elke levering bevat een pakbonbestand zodat globaal duidelijk is welke data beschikbaar zijn. Voor de vrijgave van de gegevens per 3 juli 2017 bevat het pakbonbestand (UWVopenbkaartpakbon20170703.xlsx) de volgende informatie over het databestand (UWVopenbkaartdata20170703.txt):

Categorie	Groepering	# Records	# Beroepen	# Rec. typen	# Geotypen
01 - Totalen	Peildatum: 2017-03-07 (#200)	176324	2555	3	4
02 - Recordtype	Ervaringsberoep	69874	2433	1	4
02 - Recordtype	Wensberoep	44089	2151	1	4
02 - Recordtype	Vacature	62361	1738	1	1
03 - Geotype	Postcode	96393	1741	3	1
03 - Geotype	Gemeente	52312	1195	2	1
03 - Geotype	Provincie	23035	1895	2	1
03 - Geotype	Landelijk	4584	2458	2	1
04 - Rec-/Geotype	Ervaringsberoep-Postcode	21596	189	1	1
04 - Rec-/Geotype	Ervaringsberoep-Gemeente	32513	1169	1	1
04 - Rec-/Geotype	Ervaringsberoep-Provincie	13332	1875	1	1
04 - Rec-/Geotype	Ervaringsberoep-Landelijk	2433	2433	1	1
04 - Rec-/Geotype	Wensberoep-Postcode	12436	114	1	1
04 - Rec-/Geotype	Wensberoep-Gemeente	19799	859	1	1
04 - Rec-/Geotype	Wensberoep-Provincie	9703	1520	1	1
04 - Rec-/Geotype	Wensberoep-Landelijk	2151	2151	1	1
04 - Rec-/Geotype	Vacature-Postcode	62361	1738	1	1

De eerste dataregel op de pakbon bevat informatie over de peildatum, het afslagvolgnommernummer plus 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 beroepenkaartbestand (UWVopenkaartdataYYYYMMDD.csv)

Dit bestand heeft de volgende structuur:

#	Rubrieknaam bij UWV	Betekenis
1	peildatum	Peildatum gegevens (voor alle records identiek)
2	beroep_cd	Code registerberoep UWV
3	beroepnaam_reg	Beroepnaam registerberoep UWV
4	beroep_cd_koepel	Code koepelberoep UWV
5	beroepnaam_koepel	Beroepnaam koepelberoep UWV
6	postcodegebied	Numeriek deel postcode
7	provincie	Provincienaam
8	gemeente	Gemeentenaam
9	rec_type	Record type (Ervaringsberoep, Wensberoep, Vacature)
10	aant	Totaal aantal CVs c.q. vacatures
11	leeftijd_gem	Gemiddelde leeftijd
12	aant_m	Aantal mannen
13	aant_v	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	Aantal zonder werkervaring
36	aant_we_0_1	Aantal met minder dan 1 jaar werkervaring
37	aant_we_1_3	Aantal met 1 tot 3 jaar werkervaring
38	aant_we_3_6	Aantal met 3 tot 6 jaar werkervaring
39	aant_we_7	Aantal met 7 of meer jaar werkervaring
40	aant_vcs_tot_hj	Aantal vacatures laatste half jaar



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 dichter bij MBO.