

NOTITIE

Onderwerp	Uniformering aggregatieniveau (open) datasets Energiedata
Aan	Laurens Baas
Van	Marja Exterkate (VIVET)
Datum	28 september 2020
Status	definitief

1 Samenvatting

Netbeheer NL heeft namens de netbeheerders aan Vivet de vraag gesteld op welk aggregatieniveau (ruimtelijk, temporeel) energiedata ter beschikking zou moeten worden gesteld. Aanleiding is de veelheid en diversiteit aan dataverzoeken aan de netbeheerders en de zorgvuldigheid waarmee deze verzoeken afgehandeld moeten worden. Het betreft hier open data (verbruik en opwek) op kleinverbruikersniveau.

Deze vraag wordt beantwoord in deze notitie, die geschreven is door Geonovum, als onderdeel van het Vivetproject 'Data standaarden' (project VIII.1), waar Geonovum de projectleiding over heeft. De inhoud en conclusies in deze notitie zijn vanuit de Vivet-samenwerking tot stand gekomen, aangevuld met gesprekken met experts op dit gebied en met hulp van eerder onderzoek.

Dit heeft geleid tot een set van te hanteren criteria voor de publicatie van energiedata, een opsomming van de mogelijke aggregatieniveaus en een beoordeling van de energiedata op de (vooral ruimtelijke) aggregatieniveaus. Deze notitie concentreert zich op de ruimtelijke aggregatieniveaus, omdat op de tijdschaal volledige data uitsluitend op jaar-niveau beschikbaar is.

De conclusie is dat het (CBS-)buurniveau het minimale niveau is waarop energiedata voldoet aan de gestelde eisen. Beschikbaarheid van energiedata op lagere ruimtelijke schaalniveaus wordt noodzakelijk geacht, maar waar deze al aanwezig is, voldoet deze niet aan de gestelde criteria.

Tijdens dit onderzoek is ook gekeken naar de mogelijkheid om energiedata te modelleren op gebouwniveau, waarbij de modellering is gebaseerd op zoveel kenmerken dat de resulterende waarden de statistische toets kunnen doorstaan en voldoen aan privacy-wetgeving. Deze set kunstmatige microdata wordt op regelmatige basis ververst en als open data ter beschikking gesteld, zodat de stroom aan informatieverzoeken bij de netbeheerder naar verwachting zal sterk zal inkrimpen. Niet verder uitgewerkt is de vraag welke partij welke rol speelt bij de verder de uitwerking (realisatie van de dataset, beschikbaar stellen, beheer en onderhoud, etc.).

Aanbevolen wordt om de benadering van gemodelleerde energiedata komend jaar verder uit te werken; het ligt voor de hand om dat in Vivetverband te doen.

De interviews en reacties van Vivet-projectleiders hebben mede geleid tot inzichten die indirect van belang zijn voor de vraagstelling; deze zijn opgenomen in een paragraaf 'Aandachtspunten' en in het hoofdstuk 'Aanbevelingen'.

2 Achtergrond

Netbeheer Nederland heeft aan Vivet gevraagd om een voorstel te formuleren over een bruikbaar aggregatieniveau voor open data van elektriciteit- en gasverbruik en opwek, dat zowel aan de aanbod- en de vraagkant voldoet aan de eisen en wensen van programma's en partijen die actief zijn binnen de Energietransitie (zoals de Regionale Energie Strategieën, de Transitie Visies Warmte, gemeenten, adviesbedrijven). Inzicht in het huidige verbruik/opwek is van belang voor o.a. prognoses, besluitvorming en monitoring in de context van de energietransitie.

Mede aanleiding voor dit verzoek is de grote hoeveelheid informatieverzoeken en de breedte van deze verzoeken die de netbeheerders ontvangen van diverse partijen die een rol spelen in de energietransitie.



Qua energiedata is al veel voorhanden, denk bijvoorbeeld aan alle data die noodzakelijk is om de energiemarkt te laten functioneren, maar wat ontbreekt is een helder beleid rondom (open) data om in gezamenlijkheid de grote opgave van de energietransitie goed te kunnen oppakken. Meer specifiek gaat het in deze notitie over opwek- en verbruiksdata (verder energiedata genoemd). Ook deze data zijn aanwezig maar cruciaal is het aspect in volledigheid en beschikbaarheid. Cruciaal is de combinatie en integratie van data van verschillende bronnen als open data voor algemeen gebruik. Vooral nog wordt deze integratie gehinderd door verschillen in aggregatieniveau in de verschillende databronnen, zowel in ruimte als in tijd.

De concrete vraag die voorligt:

Creëer vanuit Vivet een voorstel voor uniformering van de wijze van aggregatie bij oplevering van (open) datasets door de netbeheerders:

- Concrete afspraak voor een minimaal schaalniveau waarop opwek- en verbruiksdata geaggregeerd moeten worden;
- Werkwijze bij aanwezigheid van dominante eenheden binnen een aggregatie;
- Conform AVG, Elektriciteitswet, Gaswet en andere relevante wetgeving.

In deze notitie wordt toegewerkt naar de afspraak over dit minimaal toe te passen aggregatieniveau voor open energiedata van de netbeheerders, die dus voor ieder beschikbaar is. Er zullen tal van situaties zijn, waar data voor vereist is die niet via open data verkrijgbaar is; daarvoor dienen uiteraard de huidige informatiekkanalen.

Broninformatie

Energiedata hebben in deze context een ruimtelijke dimensie en dus een locatie. Een aansluiting, met een adres is de kleinste meeteenheid en daarmee ook broninformatie. Aggregatie van data op aansluiting-niveau is noodzakelijk vanwege privacyrichtlijnen (AVG). Aggregatie over tijd is een ander aspect. Dit heeft geen relatie met privacyrichtlijnen maar afspraken over gebruikte tijdschalen zijn van belang om eenduidige data-integratie mogelijk te maken.

Aansluitingen en adres

Energiedata wordt per aansluiting geregistreerd. Dit is meestal gekoppeld aan een adres (vbo=verblijfsobject), maar het kan ook een object zijn (tunnel, openbare verlichting etc.). De opwek (bijvoorbeeld zonnepanelen) zullen meestal ook per adres geregistreerd staan, maar er zijn natuurlijk ook zonneweides etc. Op deze problematiek wordt hier nu niet verder ingegaan. Adressen worden administratief en ruimtelijk ingedeeld middels postcodegebieden.

3 Methode

In eerste instantie zijn de voor de hand liggende en in de praktijk veelal gebruikte aggregatieniveaus beschreven in een korte interne Vivet notitie, die is voorgelegd aan de projectleiders Vivet. De in de notitie genoemde ruimtelijke aggregatieniveaus zijn de Postcode 4/6-posities en buurt/wijk/gemeente volgens de CBS-indelingen. Temporele niveaus zijn achtereenvolgens jaar, halfjaarlijks, kwartaal, maand, week, dag, uur, 15 minuten.

Het ontvangen commentaar is verwerkt en was aanleiding om enkele experts (van o.a. CBS, Netbeheer NL, PBL, Quintel en Geodan) te benaderen en het rapport 'Dataknelpunten in de energietransitie'¹ van CE Delft en Generation.Energy te betrekken dat is opgesteld in het kader van de RESsen. Daarnaast is met een schuin oog gekeken naar de nieuwe energiewet (goedkeuring medio 2022), waarin data als één van de zes pijlers een belangrijk onderwerp is.

Zoals aangegeven beperkt deze notitie zich tot open data, en qua energiedata wordt gekeken naar woningbouw (of: kleinverbruik) dus data van grootverbruikers wordt buiten beschouwing gelaten.

¹ <https://www.ce.nl/publicaties/2402/dataknelpunten-in-de-energietransitie>



4 Bevindingen

4.1 Principes en criteria

Er is een grote behoefte aan data en ook grote bereidheid om deze te delen, waarbij de afweging *inzicht versus privacy* steevast wordt benadrukt. Voor de netbeheerders is het zeer bewerkelijk dat bij heel veel informatieverzoeken deze afweging in de vorm van een privacy impact assessment gemaakt moet worden. Bij de inventarisatie van kansen en wensen kwamen daarom van nature de volgende criteria/principes aan de orde die gehanteerd zouden moeten worden bij het vaststellen van de aggregatieniveaus:

- onthullingsmogelijkheden van vertrouwelijke data (privacy, vertrouwelijkheid);
- de statistische kwaliteit van de individuele data (in CBS-termen: de microdata);
- zoveel informatie als mogelijk, bijvoorbeeld door voor samengevatte gegevens ook standaarddeviatie en/of mediaan te publiceren;
- doelondersteuning: het juiste aggregatieniveau voor de vraag;
- de mate waarin energiedata op juiste wijze te verbinden is met andere domeinen (bijv. demografie);
- de 'wendbaarheid' (veel dataverzoeken vragen in de praktijk om maatwerk);
- stabiel in de tijd: gebieden wijzigen (bijv. buurten/gemeenten) en daarom moeten historische tijdreeksen geconstrueerd kunnen worden op basis van goede verdeelsleutels;
- de optelbaarheid van energiedata tussen de verschillende gebiedsniveaus (postcode, RES, netbeheer-regio etc.);
- te bewerkstelligen tegen zo laag mogelijke lastendruk;
- alle data over het energiesysteem (elektriciteit, gas, warmte) op één plek beschikbaar en vindbaar (maar niet per sé in één systeem);
- een hartenkreet: 'kies voor een stabiele, structurele oplossing, rekening houdend met nieuwe energiewet'.

4.2 Inventarisatie van aggregatieniveaus

Temporele niveaus

In de huidige praktijk is de meeste data beschikbaar op jaarbasis. De installatie van de slimme meter heeft gezorgd voor mogelijkheden op een lager schaalniveau, tot aan 15 minutenwaarden voor elektriciteit en uurwaarden voor gasverbruik. De netbeheerders meten hun elektriciteitsnetten veelal op een nog fijnmaziger niveau (bijv. 5 minuten voor hoogspanningsnetten en delen van de middenspanningsnetten). Met het toenemende belang van flexibiliteit en integratie van de elektriciteit-, gas- en warmtesystemen stijgt de behoefte van data om op gedetailleerder tijdsniveaus van tijdsmodellering d.m.v. tijdsprofielen over te stappen op werkelijk data. Door CBS wordt aangegeven dat er vanuit Europa druk is om meer gedetailleerd dan jaarlijks te rapporteren, bv halfjaarlijks, kwartaal.

De temporele tijdsniveaus die worden genoemd zijn de voor de hand liggende (5 min / kwartier / uur / dag / week / maand / kwartaal / halfjaar / jaar).

Ruimtelijke niveaus

De volgende ruimtelijke niveaus worden in de praktijk gehanteerd:

- *Verblijfsobject* (op dat niveau wordt immers gemeten);
- *100x100 m* blokken (CBS presenteert analyses op dat niveau);
- *PC4/5/6-posities* (de Nationale Energie Atlas e.a.);
- *Buurt / wijk / gemeente / COROP/ provincie / landelijk (door het CBS);*
- *Regio:* Regionale Energie Strategie t.b.v. de RES-monitoring (door het CBS en PBL) en in de nabije toekomst door PBL voor consolidatie van regionale energiebalansen;
- De Klimaatmonitor hanteert nog de volgende aanvullende ruimtelijke niveaus:
Omgevingsdienstregio / Aanvrager VUE (informatieplicht bedrijven) / COROP-gebied / Landsdeel / Stedelijkheid, G4+G40 / Aandeelhouder . Greenport / RWS-Regio / Maatwerkregio / MIRT-regio / DuMo-regio / DeelRES Noord-Holland.

Er zijn ook wensen uitgesproken om de volgende niveaus te hanteren:

- *Bouwblok:* dezelfde type bebouwing in een gebied (in de praktijk bij Energie Atlas A'dam);
- *Straatdeel* (in de praktijk voor energiedata nog niet waargenomen);



- *Woningtype*: als aanvulling op andere ruimtelijke niveaus, bijv. ook gebieden met specifieke typen appartementen;
- *Niveau tussen de CBS-niveaus 100x100 en buurtniveau*;
- *Bedrijventerreinen*;
- *Niveaus van de net-infrastructuur, bijv. voedingsgebied van onderstation of middenspanningsruimte (transformatorhuisje)*.

De nieuwe energiewet tenslotte geeft (volgens contouren-notitie) niet direct richting aan het vraagstuk van data-aggregatie. Wel heeft de wet richtlijnen voor de wijze waarop áchter de meter gemeten wordt. Daarvoor wordt het concept van het 'secundaire allocatiepunt' benut, een meetpunt voor installaties die energie verbruiken of produceren en dat zich bevindt in het gebouw van de aangeslotene. En dat zou dan in de toekomst ook een gewenst niveau van aggregeren kunnen zijn.

4.3 Geschiktheid van niveaus

Vanuit functionaliteit gezien heeft het objectniveau (aansluiting / verblijfsobject) de voorkeur; uiteraard spelen daar de juridische beperkingen van vertrouwelijkheid en privacy, wat de reden is dat er nu (diverse) eisen zijn voor het minimaal aantal waarnemingen. Overigens zal in de nieuwe energiewet geregeld zijn, dat het de afnemer is die bepaalt wat er met zijn gegevens gebeurt, en dat opent op termijn weer andere perspectieven. Onderstaande opsomming per aggregatieniveau geeft aan welke mogelijkheden en beperkingen bestaan:

- *postcode-6*: een gedetailleerd niveau en wordt regelmatig genoemd als wens.
Bezwaren: de postcode gaat om looproutes, niet om het sommeren van energiedata; geen logische gebiedsvorm (soms samengesteld uit deelgebieden) en de bronhouder van postcodegebieden is niet helder. Vaak zijn gebieden te klein v.w.b. onthulling dan is er niet één voorgeschreven methodiek om dat op te lossen; postcodegebieden tellen niet op tot andere schaalniveaus zoals 'buurt' maar kunnen in meerdere buurten liggen;
- *postcode-5*: voornamelijk om te aggregeren vanuit postcode-6 vanwege privacy.
Bezwaren: ook postcode-5 is niet waterdicht qua onthulling; geeft vreemde gebieden en is nauwelijks te koppelen aan andere domeinen omdat postcode-5 weinig gehanteerd wordt;
- *postcode-4*: goed schaalniveau voor statistisch betrouwbare waarnemingen
Bezwaren: qua omvang vaak vergelijkbaar met buurt- of wijkniveau, dat dan de voorkeur heeft;
- *100x100 m*: in sommige gebieden goed toe te passen (bijv. bij dichte gelijkvormige bebouwing)
Bezwaren: dit zijn geen herkenbare vlakken; zijn niet overal toe te passen vanwege soms geringe aantallen waarnemingen; in de praktijk niet toegepast voor waargenomen datawaardes, maar wel voor afgeleide data (bijv. percentages);
- *buurt- en wijkniveau*: veel toegepast, gemakkelijk te combineren met andere domeinen; goed optelbaar, gemeente bepaalt indeling waarmee logische indelingen mogelijk zijn
Bezwaren: ook buurtniveau kan te laag zijn i.v.m. onthulling;
- *buurt / wijk / gemeente*: standaard-indeling, veel toegepast
Bezwaren: niet stabiel in de tijd vanwege inter- en intragemeentelijke herindelingen;
- *RES-niveau*: een nieuw niveau, goed werkbaar omdat RES-niveau een aggregaat is van gemeenten en goed aansluit bij intensieve bestuurlijke samenwerking op energiegebied
Bezwaren: dit niveau zal zich nog moeten bewijzen;
- *bouwblok, straatdeel*: door de selectie van kleinschaliger gebieden in de openbare ruimte zou het mogelijk kunnen zijn om gelijkvormige bebouwing te onderkennen, wat interessant is voor energiedata
Bezwaren: er is geen standaard (diverse algoritmen); weinig toegepast in de praktijk, geen eenduidige norm; ook hier kans op onthulling; niet bestendig in de tijd (sloop, nieuwbouw);
- *woningtype binnen een buurt/wijk*: gelijkvormige woningen kunnen interessant zijn voor analyse van energiedata. Selecties binnen hogere schaalniveaus kunnen voldoen aan onthullingscriteria (bijv. twee-onder-een-kap per buurt of per gemeente)
Bezwaren: dit is waarschijnlijk ook niet generiek toe te passen op een aantrekkelijk schaalniveau: er zullen altijd buurten zijn met een zeer beperkt aantal woningen van één woningtype.
- *bedrijventerreinen*: interessant omdat in de praktijk ondernemers soms gezamenlijk op willen trekken en qua energie synergie kunnen bereiken door samen te werken;
Bezwaren: als open data lastig vanwege onthulling en door grote variatie statistisch niet goed werkbaar;



- *niveaus van netinfrastructuur* (onderstations, transformatorhuisjes): het net moet de flexibilisering van de energieproductie (wind, pv) accommoderen. De beschikbaarheid van netcapaciteit is in toenemende mate van belang voor projectontwikkelaars en andere betrokkenen bij de planning van nieuwe productiesites.
Bezwaren: verschillende benaderingen door de netbeheerders; gevaar voor onthulling (bijv. bij grote aangeslotenen op het net);
- *De hierboven niet vermelde in de Klimaatmonitor toegepaste ruimtelijke niveaus*: geschikt voor specifieke doeleinden.

Bovenstaande bevindingen leiden tot de conclusie, dat er - vooral voor de gedetailleerdere ruimtelijke aggregatieniveaus - veel bezwaren zijn, terwijl er juist veel behoefte is aan data op kleinere ruimtelijke schaal. Niet heel verwonderlijk, maar dat maakt het lastig om een acceptabel aggregatieniveau te adviseren. Daarom wordt hieronder in kort bestek een alternatieve benadering beschreven die analoog is aan de methode die gebruikt wordt voor de Startanalyse van PBL.

4.4 Gemodelleerde energiedata

De Startanalyse rekent met gemodelleerde verbruiken per woning, op basis van statistisch verantwoorde gemiddelden van vergelijkbare woningtype, bouwperiode, vloeroppervlakteklasse en energielabel. Deze gemiddelden zijn door CBS bepaald op basis van de werkelijke energiedata, en die zijn vastgelegd in een tabel met de waarden voor enkele duizenden combinaties. Om dichterbij de werkelijke waarden te komen, vinden nog correcties plaats voor seizoenstemperatuur en warmtapwater.

Op deze manier ontstaan fictieve waarden (kentallen) die toch een goed beeld geven van de werkelijke energiedata. Afwijkingen kunnen veroorzaakt worden door diverse omstandigheden: ruimtelijk, meteo, demografie, energiegedrag.

Deze afwijkingen zullen deels gecorrigeerd kunnen worden door deze te ijken op het buurttotaal t.o.v. landelijke gemiddelden.

Deze kentallen kunnen vervolgens worden toegepast op het kleinverbruikersbestand en zo ontstaat een kunstmatige microdataset. Deze biedt de mogelijkheid om vervolgens op elk gewenst aggregatieniveau naar wens groepen te gaan vormen.

Met inachtnaam van de beperking dat dit gemodelleerde gegevens zijn, komt deze aanpak tegemoet aan de behoefte aan data op fijnmazig ruimtelijk aggregatieniveaus.

Er zullen naar verwachting nog methodologische haken en ogen zijn, die onderzocht moeten worden. Te denken is hierbij aan o.a. toetsen op statistische kwaliteit en privacybescherming.

4.5 Aandachtspunten

Op basis van feedback op de eerste conceptversie van deze notitie en recente ontwikkelingen volgen hier nog enkele aandachtspunten die direct of indirect te maken hebben met de keuze voor aggregatieniveaus van open energiedata.

- de behoefte aan open data zou je nog breder in beeld kunnen brengen (meer vraaggestuurd dus), om op die wijze open data zo breed en eenvoudig mogelijke beschikbaar te maken;
- benut hierbij de expertise van CBS en andere Vivet-partijen;
- zorg dat energiedata van gebouwde omgeving zo goed mogelijk te relateren is aan andere data van het energiesysteem, zoals andere verbruikers (utiliteitsbouw, openbare verlichting e.d.), netwerkdata, energie-installaties;
- let op bij publicatie van open data op de synchroniteit tussen tijd van meting en tijd van publicatie;
- een roep vanuit gebruikers: voorlopige cijfers zijn soms wenselijk in verband met de behoefte aan actuele data; goed versiebeheer bij het actualiseren is noodzakelijk;
- de vraag wordt gesteld of het verschil maakt of (en of het juridisch toegestaan is dat) de bronhouder (hier: netbeheerder) zijn 'kale' data publiceert, of dat deze de data zelf ook verkrijgt;



5 Conclusie

Ruimtelijke schaal

Het minimale aggregatieniveau van werkelijke energiedata dat goed toegepast kan worden, is CBS-buurtniveau. De beschikbaarheid op een ruimtelijk niveau lager dan CBS-buurtniveau is zeer gewenst maar er is geen lager niveau dat vrij is van praktische en juridische bezwaren.

Om tegemoet te komen aan de wens van energiedata op gedetailleerder schaalniveau, kan mogelijk op termijn gemodelleerde data in een kunstmatige microdataset beschikbaar worden gesteld, waarmee naar behoefte sets van energiedata op elk aggregatieniveau kunnen worden samengesteld.

Temporele schaal

Volledige datasets zijn voorlopig nog vooral op jaarbasis. Met de implementatie van de slimme meter komt steeds meer data op lager schaalniveau beschikbaar. De nieuwe energiewet specificeert rollen voor het beschikbaar maken van open data op een lager schaalniveau. De netbeheerders beschikken over fijnmazige (per 5 minuten) data over hun elektriciteitsnetten. Het al dan niet ter beschikking stellen daarvan en de wijze waarop zal onderwerp van discussie blijven.

6 Aanbevelingen

- De benadering van gemodelleerde energiedata wordt door geïnterviewden als kansrijk gezien; wanneer deze gemodelleerde data op objectniveau (verblijfsobject) wordt toegewezen en als open data beschikbaar wordt gesteld, kunnen gebruikers naar wens een aggregatie voor hun toepassing maken. Naar verwachting zorgt het voor een geringer belasting bij de netbeheerders.
- Onderzoek op korte termijn of het mogelijk is om een minimaal kwaliteitsniveau te behalen, wat de juiste wijze is om deze data ter beschikking te stellen en wie welke rol moet vervullen. De basiselementen om deze te genereren zijn aanwezig (CBS, Startanalyse PBL). Hier zijn rollen voor CBS, PBL, Netbeheer NL; uitwerking kan plaatsvinden onder de Vivet-paraplu. Betrek hierbij naast overheden ook vertegenwoordigers van woningcoöperaties en adviseurs (bijv. EG-ETRM, de Expert Groep Energie Transitie Rekenmodellen).
- Zorg dat deze data ook zo bruikbaar mogelijk ter beschikking wordt gesteld en dat dit voor de lange termijn geregeld wordt.
- Hier speelt naast eenvoudige toegang ook actualiteit een rol (frequent – maandelijks? - actualiseren en een snelle procedure om de data te modelleren).
- Waar gewenst zou geanticipeerd kunnen worden op het analoog beschikbaar stellen van gegevens over warmtenetten.
- Anticipeer op de nieuwe energiewet, waar de energie-datavoorziening één van de pijlers is, zoals beschreven in de nota Contouren van de Energiewet².