

Datavalidatie

Er is een steeds grotere vraag naar historische data en deze worden voor vele doeleinden gebruikt. Het maakt niet uit waar data voor gebruikt worden; feit is dat de data alleen waarde hebben als het goede data zijn. Goede data zijn data die overeenkomen met de werkelijkheid op het moment van registreren. Via het TMX-telemetriesysteem van de afdeling Beheer Watersystemen & Wegen binnen Hoogheemraadschap Hollands Noorderkwartier (HHNK), genaamd BOA, worden vele data verzameld en verwerkt. De BOA verwerkt binnengekomen data van peilregulerende kunstwerken en diverse meetpunten.

VAN PEILSCHAAL TOT DATABASE

‘Meten is weten’, maar wat weten we nu eigenlijk als we gemeten data ontvangen. Als uitgangspunt wordt aangehouden: een goede meetwaarde is een waarde die op het tijdstip van registratie overeenkomt met de werkelijkheid.

Data die op de BOA binnenkomen, zijn waarden die in het verleden liggen en dus niet meer te vergelijken zijn met de werkelijkheid. Er moet een validatie uitgevoerd worden om te controleren of de ontvangen data realistisch zijn ten opzichte van de historie. Indien aan de validatiecriteria wordt voldaan, kan worden gesteld dat er beschikking is over een goede meetwaarde. Niet alle data hoeven per definitie gevalideerd te worden. Alleen die data die nodig zijn voor latere bewerkingen komen in aanmerking als historische data. Denk hierbij bijvoorbeeld aan peilmetingen, debieten en stuwstanden, maar een accuspanningmeting zal niet gevalideerd hoeven worden.



Bij een peilregulerend kunstwerk zijn er diverse parameters die worden gemeten en geregistreerd. Bij iedere meting zijn er diverse aandachtspunten om tot een zo nauwkeurig mogelijke meting te komen. Er wordt hierbij onderscheid gemaakt tussen dynamische en statische opnemers. Een dynamische opnemer is een opnemer waarin zich onderdelen bevinden die bewegen

om een meting te kunnen verrichten. Een duidelijk voorbeeld hiervan is een windmeter die draait als het waait en zo meetwaarden genereert.

Een statische opnemer is een opnemer waarin zich geen bewegende onderdelen bevinden. Te denken valt hierbij aan een signaal uit een frequentieomvormer of een temperatuurmeting. Deze parameters dienen bij aanvang goed te worden ingeregeld en behoeven in principe geen nazorg meer te hebben. Bij dynamische opnemers dient rekening te worden gehouden met slijtage en dus een verloop van de meting.

PEILMETING

De meest voorkomende meting bij peilregulerende kunstwerken is de peilmeting. Voor het verkrijgen van een goede meting is er een aantal stappen te doorlopen. Iedere stap dient zo nauwkeurig mogelijk te worden uitgevoerd, omdat iedere afwijking doorwerkt in het gehele traject tot het genereren van de feitelijke meetwaarden.

Het eerste aandachtspunt is de opnemer zelf. De opnemer dient een zodanige nauwkeurigheid te hebben dat een inwendige meetfout door de opnemer zelf geminimaliseerd wordt. De opnemer zal vervolgens gekalibreerd moeten worden, dat wil zeggen dat het gemeten signaal omgezet moet worden naar leesbare data.

De kalibratie geschiedt door het gemeten signaal gelijk te zetten aan de afgelezen waarde op een ter plekke aanwezige peilschaal. Hier dient het tweede aandachtspunt aan. De peilschaal moet zodanig zijn ingemeten en geplaatst dat de peilschaal geen fout aangeeft. Uiteindelijk komt het aflezen van de peilschaal zelf, dit dient met aandacht te gebeuren om een zo nauwkeurig mogelijke kalibratie te verkrijgen.

Ten behoeve van de controle tussen peilschaal, opnemer en telemetrieregistratie wordt gebruik gemaakt van een PDA en een standaard invulformulier, dat via e-mail direct verzonden kan worden. Door deze werkwijze zijn papierstromen verminderd, is de verwerkingsnelheid toegenomen en worden de telemetriebeheerders in hun werk minder gestoord door telefoontjes. Het uitgangspunt is dat met een interval van twee maanden de opnemers gecontroleerd worden.

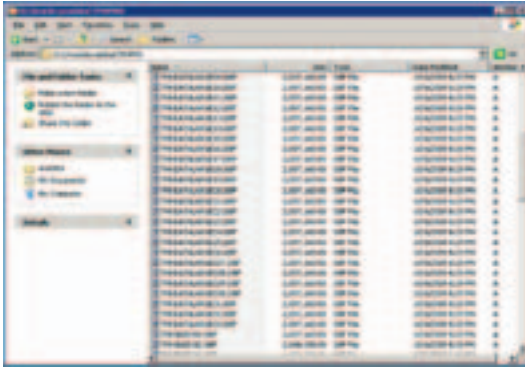
Formulier voor peilmeting met de volgende secties:

- Algemeen: Naam, Adres, Telefoon, E-mail, Datum, Plaats, etc.
- Peilmeting: Tabel met 6 kolommen (Meting 1 t/m 6) en 1 rij voor de gemiddelde waarde.
- Meting peilschaal (mLSP): Tabel met 6 kolommen (Meting 1 t/m 6) en 1 rij voor de gemiddelde waarde.
- Meting telemetrie (mLSP): Tabel met 6 kolommen (Meting 1 t/m 6) en 1 rij voor de gemiddelde waarde.

Uit het voorgaande blijkt dat er diverse factoren aanwezig zijn die van invloed zijn op het juist verkrijgen van data. Een steeds teugkerend punt van aandacht is onderhoud. Afhankelijk van de organisatiestructuur kan het voorkomen dat de ene afdeling onderhoud uitvoert op een kunstwerk waar de andere afdeling, verantwoordelijk voor meetdata, hier geen weet van heeft, maar wel de gevolgen van meekrijgt. Als er bijvoorbeeld onderhoud uitgevoerd wordt aan een neerslagmeter dan kan er ter controle water in worden gegoten om te zien of alle componenten nog werken. De onderhoudsmedewerker heeft zijn taak uitgevoerd en gaat naar de volgende opdracht. Echter wat hij zich vaak niet realiseert is dat hij, door het gieten van water, meetwaarden heeft gegenereerd.

Deze data zijn niet realistisch omdat die niet overeenkomt met de werkelijkheid. Het zou zo moeten zijn dat als er werkzaamheden uitgevoerd worden op een kunstwerk, dit gemeld wordt aan de databeheerder zodat deze alle niet realistische data weg kan werken.

Als je dit zo leest denk je bij je zelf 'logisch toch dat je dit even meldt', maar dat is theorie. In de praktijk blijkt dat dit nogal eens wordt vergeten. Het is



een cliché maar het terugdringen van deze meetfouten staat of valt met **communicatie**. Als databeheerder wordt hier dan ook veel extra aandacht aan besteed.

VALIDATIE

Voordat je overgaat tot validatie moet je je afvragen of het noodzakelijk is om alle parameters te valideren. De eerste

stap is dus het bepalen welke parameters van belang zijn om voor validatie in aanmerking te komen. Gesteld kan worden dat alleen meetgegevens, die naar de lange termijn database overgezet gaan worden, worden gevalideerd. In het geval van peilregulerende kunstwerken zijn dit: peilmetingen – debietmetingen en -berekeningen – frequentiemetingen – stuwstanden – inlaatstanden – in/uitbedrijfmeldingen.

Validatie heeft tot doel binnengekomen data te controleren op onwaarheden. Er kan niet uitgegaan worden van generieke waar- en onwaarheden. Als voorbeeld: er kan niet worden gesteld dat een meetwaarde binnen een vaste tijdsinterval altijd een veranderende meetwaarde geeft. Een neerslagmeter geeft namelijk alleen data als het regent. Er zijn per parameter of groep van soortgelijke parameters validatieregels opgesteld.

Binnen de BOA beschikken we over primaire en secundaire validatiemogelijkheden. Primaire validatie zijn generiek toepasbare regels. Dit is een automatisch proces waarbij binnenkomende data direct van kwaliteitskenmerken worden voorzien. Een voordeel van validatiemogelijkheden binnen de BOA is dat validatie dicht bij de bron, direct bij binnenkomst van de data plaatsvindt.

Om validatie te verfijnen kan er gebruik gemaakt worden van niet-basisregels, de zogenaamde secundaire validatie. Secundaire datavalidatie is het proces waarbij in de database aanwezige data, achteraf worden voorzien van een kwaliteitskenmerk. Ook kunnen missende data worden aangevuld en eerdere validatiekenmerken aangepast worden.

Secundaire validatie is in hoofdlijnen onder te verdelen in twee mogelijkheden: handmatig of automatisch. Handmatige secundair validatie is het in de database per registratie aanpassen van kenmerken. Automatische secundaire validatie is complexer, maar biedt wel meerdere mogelijkheden. Bij secundaire automatische validatie worden er vergelijkingen gemaakt met verkregen meetdata van zowel eigen kunstwerk(en) als metingen verkregen van externe bronnen.

Een aantal voorbeelden van secundaire validatieregels.

- Een vergelijking van een meting achter het krooshek en instroom, heeft als voorwaarde dat een instroommeting nooit lager is dan de meting achter het rooster;

- Als een gemaal niet draait, moet de meting instroom en achter het rooster gelijk zijn aan elkaar;
- Bij een inlaat kan de benedenstroomsmeting nooit hoger zijn dan de bovenstroomsmeting;
- Bij een stuw en inlaat kan er een vergelijking tussen de peilmetingen worden gemaakt met een boven- dan wel een benedenstrooms geplaatste stuw, gemaal of inlaat, waar zich ook opnemers bevinden;
- De meting aan de uitstroomzijde kan worden vergeleken met uitstroommetingen van naastgelegen kunstwerken, die op dezelfde waterloop zijn aangesloten;
- Indien er een inbedrijf-melding is kan de frequentiemeting geen nul zijn;
- Debietmetingen kunnen pas een waarde geven indien er een inbedrijf-melding is bij een gemaal;
- Debietmetingen kunnen pas een waarde geven indien een klep openstaat bij een spuilocatie;
- Debietmetingen kunnen pas een waarde geven indien klepstand lager is dan het bovenstrooms peil.



Om te kunnen valideren dienen er uiteraard wel data aanwezig te zijn. Dit kan op twee manieren. De database openen en alle data controleren op aanwezigheid. Maar bij een omvangrijk systeem als bij HHNK, met meer dan 250.000 registraties per dag, is dit niet meer werkbaar. Door gebruik te maken van scripting is het mogelijk om heel gericht te zoeken naar ontbrekende data. Zo wordt automatisch een terugmelding gegenereerd van welke meting, kunstwerk en tijdstip er data ontbreken. Door nu secundaire validatie toe te passen kunnen de ontbrekende data alsnog worden toegevoegd om een gesloten meetreeks te verkrijgen.

De scriptingmogelijkheden worden ook gebruikt om data te achterhalen die niet zijn goedgekeurd. Hierop kunnen door een gegevensbeheerder dan, indien nodig, correcties worden uitgevoerd.

Inmiddels is er al veel ervaring op gedaan met validatie in de praktijk. Onze conclusie is dat valideren niet iets is van 'even erbij doen'. Het is veel meer dan even een datareeks bekijken. Ondanks dat er veel geautomatiseerd kan worden, hoe creatief je ook bent in het gebruikmaken van scripts, er blijft veel nazorg nodig van een databeheerder om gesloten datareeksen met juiste data te verkrijgen.

*Arjen Immink,
Hoogheemraadschap Hollands Noorderkwartier*