

Datum: 12 maart 2013
Aan: Gebruikers en beheerders Bodeminformatiesystemen
Onderwerp: Notitie methoden voor data-analyse bodeminformatie

Introductie

Voor optimaal gebruik van bodeminformatie is het van belang dat de geregistreerde informatie voor zover mogelijk correct, uniform en consistent (samenhangend) is. Dit is ook in het belang bij het beheer van bodeminformatie. Om een indruk te krijgen van de kwaliteit (juistheid, uniformiteit en consistentie) van de geregistreerde informatie, kan een data-analyse worden uitgevoerd. Een dergelijke data-analyse wordt ook wel datacontrole of kwaliteitscontrole genoemd.

Voor het uitvoeren van een data-analyse is het noodzakelijk dat eerst inzichtelijk wordt gemaakt welke eisen door de organisatie die bodeminformatie beheert stelt aan de kwaliteit van deze informatie. Dit kunnen wettelijke eisen voor bodeminformatie zijn, maar ook eisen die de opdrachtgever zelf stelt aan de geregistreerde bodeminformatie bijvoorbeeld vanwege gebruik van informatie voor eigen bedrijfsprocessen. De eisen kunnen zijn vastgelegd in bijvoorbeeld een invoerprotocol. Aan de hand van de inventarisatie van de kwaliteitseisen worden controlecriteria opgesteld. De controlecriteria kunnen worden vertaald naar SQL-scripts waarmee de geregistreerde bodeminformatie geautomatiseerd wordt bevraagd. Naast het controleren van geregistreerde informatie met SQL-scripts kan ook een handmatige controle worden uitgevoerd. Hierbij wordt de geregistreerde informatie vergeleken met de informatie die in de bodemdossiers aanwezig is. De controle kan worden uitgevoerd op (een steekproef van) alle geregistreerde gegevens of van een selectie van de geregistreerde gegevens op basis van specifieke selectiecriteria.

Verskillende methoden voor controle kwaliteit bodeminformatie

De controle van de kwaliteit van bodeminformatie kan op verschillende manieren worden uitgevoerd. Kenmerkend voor een controle is dat vooraf een set criteria is vastgesteld op basis waarvan de bodeminformatie wordt gecontroleerd. Aangezien het grootste deel van de bodeminformatiesystemen de informatie vastleggen in een database (meestal Oracle of SQL-server) is een veel gebruikte methode voor de controle het uitvoeren van SQL-commando's. Een andere veelgebruikte methode is de controle van bodeminformatie die is vastgelegd in XML-bestanden volgens het SIKB-protocol 0101 met de zogenaamde LIB-controletool (www.sikb0101-libcontroletool.nl). Een andere mogelijkheid is om (een deel van) de bodeminformatie uit het bodeminformatiesysteem te exporteren naar Excel-bestanden en de informatie verder in Excel te analyseren. Verder is het ook mogelijk om gegevens handmatig te controleren, wat dikwijls steekproefsgewijs wordt gedaan.

In de hierna volgende tekst worden deze methoden verder toegelicht.

Controle met LIB-controletool

De LIB-controletool is bedoeld voor de controle van gegevens die zijn vastgelegd in een XML-bestand dat is ingedeeld volgens het SIKB-protocol 0101 (kortweg SIKB0101). Het doel van de controle is "*om de kwaliteit van de gegevens die overheden invoeren en uitwisselen met elkaar te verbeteren*".¹ De controle van de LIB-controletool beperkt zich tot de zogenaamde LIB-dataset. Dit is een deelset van het SIKB0101 en omvat alle gegevens die relevant zijn voor uitwisseling van bodeminformatie tussen overheden die bevoegd gezag Wbb zijn. De LIB-dataset omvat gegevens die nodig zijn voor de monitoring bodemsanering/MTR en voor het landelijke bodemloket (bodemloket.nl). De LIB-dataset bevat voornamelijk algemene gegevens over locaties, onderzoeken, verontreinigings- en saneringscontouren, kosten en besluiten. Detailgegevens zoals meetpunten, analysemonsters en analyseresultaten zijn nadrukkelijk geen onderdeel van de LIB-dataset; deze worden dan ook niet met de LIB-dataset gecontroleerd.

De LIB-controletool kent een aantal deelsets waarvoor specifieke controleregels gelden. Deze deelsets zijn: MTR en Bodemloket.

De LIB-controletool maakt gebruik van controleregels die via de website zijn te downloaden.

Het resultaat van een controle met de LIB-controletool is een CSV-bestand met geconstateerde "ERRORS" en "WARNINGS" en een samenvatting. Errors zijn fouten die conform de acceptatieregels voor uitwisseling van bodeminformatie opgelost moeten worden. Deze fouten betreffen voornamelijk het ontbreken van belangrijke informatie. Warnings worden verdeeld in algemene warnings en warnings per deelset. Algemene warnings betreffen mogelijke onjuistheden in de samenhang (consistentie) van de informatie. Warnings per

¹ Citaat van www.sikb0101-libcontroletool.nl

deelset houden verband met het mogelijk voorkomen van onvolledigheden of onsamenhangende informatie die sterk afhankelijk zijn van de fase van onderzoek/sanering die op de locatie van toepassing is.

Om een controle met de LIB-controletool uit te kunnen voeren, wordt een export gemaakt van de gegevens van de locaties en daaraan gekoppelde informatie die gecontroleerd moet worden. Als een grote hoeveelheid locaties gecontroleerd moet worden, is het zinvol een SIKB-exportbestand te maken waarin geen detailgegevens zoals meetpunten en analysemonsters zijn opgenomen. Door deze gegevens uit te sluiten bij de export wordt voorkomen dat het SIKB-exportbestand erg groot wordt.

LET OP: Zoals eerder is aangegeven controleert de LIB-controletool alleen informatie in velden die voorkomen in de LIB-dataset. Wanneer naast de velden van de LIB-dataset ook andere velden gecontroleerd moeten worden, dan is een controle die enkel is uitgevoerd met de LIB-controletool niet afdoende.

Controle met SQL-commando's

Bij de meeste bodeminformatiesystemen wordt de informatie opgeslagen in een database. Meestal is dat een Oracle-database, maar ook andere database-software wordt gebruikt zoals SQL-server en MS Access. Om de informatie die in een database is opgeslagen inzichtelijk te maken wordt de computertaal SQL gebruikt. SQL staat voor Structured Query Language. Door het uitvoeren van een SQL-commando kan informatie uit de database worden verzameld en in een overzicht getoond. Door in het SQL-commando te verschillende criteria op te nemen, kunnen zeer gerichte overzichten van informatie worden getoond. Met behulp van SQL-commando's is het mogelijk om elke denkbare zoekvraag en daarmee elke denkbare controle vorm te geven en uit te voeren. Met SQL-commando's kan worden gekeken naar ontbrekende informatie (lege velden) en naar de samenhang tussen verschillende velden zowel binnen één niveau (bijvoorbeeld binnen een locatie) als tussen verschillende niveaus (bijvoorbeeld tussen een locatie en een onderzoek).

Indien gewenst is het zelfs mogelijk om in een controle met SQL-commando's ook informatie van buiten de database te betrekken zoals geometrie. Hiervoor is het overigens wel noodzakelijk dat deze informatie bereikbaar is voor programma waarmee het SQL-commando wordt uitgevoerd.

De volgende gegevens kunnen met SQL-commando's worden gecontroleerd (lijst is niet uitputtend):

- Locatie:
 - Volledigheid adresgegevens:
 - ontbrekende straatnamen, huisnummers, postcodes, plaatsnaam
 - correcte relatie tussen straatnaam, huisnummer, postcode en plaatsnaam.
 - Vulling statusvelden:
 - volledigheid velden (controle op lege velden);
 - correcte relatie tussen statusvelden onderling en tussen statusvelden en gerelateerde informatie zoals onderzoeken, besluiten, onderzochte activiteiten (bijvoorbeeld: er is wel een besluit instemming saneringsplan maar geen besluit instemming uitgevoerde sanering terwijl het WBB-vervolg is gevuld met 'voldoende gesaneerd').
 - Volledigheid overige detailvelden.
 - Koppeling met onderzoeken, besluiten, verontreinigings- en sanerings- en nazorgcontouren, kosten en onderzochte activiteiten.
- Onderzoek:
 - Volledigheid adresgegevens:
 - ontbrekende straatnamen, huisnummers, postcodes, plaatsnaam;
 - correcte relatie tussen straatnaam, huisnummer, postcode en plaatsnaam.
 - Volledigheid onderzoeksgegevens:
 - juistheid type onderzoek, aanleiding en eindoordeel/vervolg van het onderzoek in relatie tot elkaar en overige gegevens;
 - ontbrekende gegevens zoals rapportdatum, rapportauteur, rapportnummer, conclusie.
 - Volledigheid gegevens meetpunten, analysemonsters, analyseresultaten:
 - ontbrekende gegevens zoals diepte, matrix, soort, resultaat;
 - onderlinge relatie gegevens zoals overeenkomst diepte meetpunt met diepte analysemonster;
 - juistheid gegevens ten aanzien van soort, grootte, enz.
 - Koppeling met locatie.
- Contouren:
 - Volledigheid gegevens aard en omvang.
 - Koppeling met locatie.
- Besluiten:
 - Volledigheid gegevens: datum, type besluit, kenmerk, status.
 - Koppeling met locatie.

Analyse bodeminformatie met behulp van Excel

Het is mogelijk om bodeminformatie te analyseren met behulp van Excel. Hiervoor is het nodig dat de informatie die geanalyseerd moet worden, wordt geëxporteerd vanuit de database naar een Excel-bestand. In het Excel-bestand kan de informatie worden geanalyseerd door filters toe te passen of draaitabellen te maken. Wanneer het de bedoeling is om met Excel informatie uit verschillende tabellen van het bodeminformatiesysteem te analyseren dan kan dit niet zonder gebruikt te maken van SQL. Analyse van bodeminformatie met behulp van Excel kan een goede aanvulling zijn op de controle van bodeminformatie met behulp van SQL-commando's. Door uitkomsten van SQL-commando's verder te analyseren met Excel, kunnen deze uitkomsten verder gegroepeerd en uitgewerkt worden.

Handmatige controle

Bodeminformatie kan handmatig worden gecontroleerd. Gelet op de hoeveelheid informatie die gewoonlijk in bodeminformatiesystemen is opgeslagen zal een uitvoeren van een handmatige controle van alle informatie veel tijd in beslag nemen. Het uitvoeren van een steekproefsgewijze handmatige controle vraagt aanzienlijk minder tijd, maar hiermee kunnen veel fouten in de database onopgemerkt blijven. Het uitvoeren van een handmatige controle in combinatie met een controle met SQL-commando's is beduidend effectiever. Met de controle met SQL-commando's kunnen gericht locaties en onderzoeken worden geselecteerd waarbij mogelijk onvolkomenheden voorkomen. Middels een handmatige controle van deze locaties kan de informatie in het bodeminformatiesysteem worden vergeleken met de informatie in het bijbehorende dossier. Geconstateerde fouten kunnen direct worden verbeterd.

Verbetering van geconstateerde 'fouten'

De bij een data-analyse geconstateerde onvolkomenheden (gegevens zijn niet juist, uniform of consistent) kunnen handmatig dan wel geautomatiseerd worden verbeterd. Bij handmatige verbetering wordt de entiteit (locatie of onderzoek) waarbij de onvolkomenheid is geconstateerd opgezocht in het bodeminformatiesysteem en wordt de informatie verbeterd. Wanneer informatie geautomatiseerd verbeterd kan worden, worden gegevens in de database verbeterd met behulp van SQL-commando's.