

KWR 2016.062 | 2016

HydroMonitor - open data exchange format (toelichting en definitie)

KWR
Watercycle
Research
Institute

HydroMonitor - open data exchange format (toelichting en definitie)

KWR 2016.062 | 2016

Project manager
Edu Dorland

Kwaliteitsborger(s)
E. Vonk

Auteur(s)
J.R. von Asmuth, E. Vonk

Verzonden aan
Vrij beschikbaar op internet via onderstaande link:

http://www.kwrwater.nl/downloads/menyanthes/hydromonitor_data_exchange_format.pdf

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

Watercycle Research Institute

KWR | 2016 © KWR

All rights reserved.
No part of this publication may be reproduced, stored in an automatic database, or transmitted, in any form or by any means, be it electronic, mechanical, by photocopying, recording, or in any other manner, without the prior written permission of the publisher.

Contents

Contents 2

1 Inleiding 3
1.1 Achtergrond en aanleiding 3
1.2 Kenmerken van het HydroMonitor format 3
1.3 Versiebeheer en historie 4
1.4 Opmerkingen, suggesties en aanvullende informatie 4

2 Principes en uitgangspunten 5
2.1 Identificatie, data en metadata van een ‘dataobject’ 5
2.2 Objecttypen of HydroObjects van de HydroMonitor 5

3 Beschikbare varianten 6
3.1 Spreadsheet- of matrixvariant (CSV en Excel, niet-hiërarchisch) 6
3.2 Binaire variant (MEF, hiërarchisch) 6

4 Gegevensdefinities 8
4.1 Format, datatype, eenheid en/of referentie 8
4.2 Betekenis van velden en kolommen 9

Bijlage A: Het HydroMonitor Excel format 11

Bijlage B: Het HydroMonitor CSV format 12

Bijlage C: Het HydroMonitor MEF format 13

1 Inleiding

1.1 Achtergrond en aanleiding

KWR investeert en voert verbeteringen door in haar programma's Menyanthes en HyCA voor respectievelijk waterkwantiteit en waterkwaliteit, als onderdeel van de migratie van beide naar een nieuw, geïntegreerd programma voor hydrologische gegevens, onder de naam HydroMonitor. De huidige functionaliteit van Menyanthes en HyCA zal daarin terugkomen als modules HydroTimes, HydroData en HydroChemistry. De afgelopen periode is veel aandacht besteed aan het toevoegen en verbeteren van de functionaliteit die betrekking heeft op 'data' in brede zin. Het gaat daarbij bijvoorbeeld om databeheer en datakwaliteitscontrole (QC), maar ook om het stroomlijnen van data-inwinning in het veld, en het verbeteren en vereenvoudigen van data-uitwisseling.

Voor het verbeteren en vereenvoudigen van de data-uitwisseling is de import- en exportfunctionaliteit van Menyanthes geheel opnieuw ontworpen en vervangen. Als onderdeel daarvan is een nieuw, HydroMonitor data-uitwisselformat in het leven geroepen met daaraan gekoppeld een nieuw, onderliggend datamodel. Naast KWR heeft ook Waternet bijgedragen aan de totstandkoming van dit format.

We lichten in de document de opzet, kenmerken en achtergronden van het nieuwe format toe, en specificeren de bijbehorende datadefinities. Dit document is vrijelijk en open beschikbaar op internet via onderstaande link:

http://www.kwrwater.nl/downloads/menyanthes/hydromonitor_data_exchange_format.pdf

Op die manier kunnen zowel u als anderen daar kennis van nemen, en hun voordeel mee doen.

1.2 Kenmerken van het HydroMonitor format

Ondanks de lopende initiatieven voor het ontwikkelen van open standaarden voor data-opslag en data-uitwisseling, ontkomt een specialistisch programma als Menyanthes nog niet aan het hanteren van een eigen format voor data-uitwisseling. Menyanthes heeft, of liever gezegd had, sinds jaar en dag een eigen bestandsformat, dat losjes gebaseerd was op de CSV-formats van de DINO database van TNO. Dit format kent echter een aantal beperkingen en nadelen, reden waarom nieuwe versies van Menyanthes het niet langer ondersteunen. In toekomstige versies zal Menyanthes het nieuwe HydroMonitor format voor data-uitwisseling gebruiken. Dit format heeft op hoofdlijnen als kenmerken:

- Interne definitie van het format (naam, versie, type, inhoud)
- Interne definitie van variaties in het format (datatype, data-inhoud, objecttype, locatie-ID label of sleutel)
- Interne definitie van kolommen of velden (datatype, eenheid, referentie)
- Aanvullende definities en toelichting zijn open en beschikbaar via de gegeven link
- Beschikbaar in Excel-, CSV – en MATLAB-variant
- Ondersteund bij zowel import als export

1.3 Versiehistorie

Het HydroMonitor format is voorzien van versienummering en versiebeheer. De verschenen versies zijn:

Versie	Datum release	Omschrijving
1.0	2016-09-30	Eerste openbare release van het HydroMonitor format. Deze versie zal in toekomstige versies ondersteund blijven worden.
0.1	Geen	Ontwikkelversie, niet gereleased maar verspreid onder een selecte groep geïnteresseerden.

1.4 Opmerkingen, suggesties en aanvullende informatie

Het HydroMonitor format staat open voor verbetering. Heeft u opmerkingen, suggesties of behoefte aan aanvullende informatie? Neem dan contact op met Jos von Asmuth (Jos.von.Asmuth@kwrwater.nl, 030-6069512) of Erwin Vonk (Erwin.Vonk@kwrwater.nl, 030-6069547).

2 Principes en uitgangspunten

2.1 Identificatie, data en metadata van een 'dataobject'

In de opzet van het HydroMonitor format speelt de notie van een 'dataobject' een centrale rol. Met de term object refereren we hier aan een uniek 'iets' waarover gegevens voorhanden zijn, zoals een peilbuis, weerstation of een datalogger. Op de achtergrond speelt hierbij mee dat binnen de HydroMonitor verschillende typen objecten of 'HydroObjects' worden onderscheiden, waarover later meer. Op het hoogste niveau van abstractie heeft een object de volgende vier typen gegevens of eigenschappen:

- **Type**: De klasse of het objecttype waartoe een individueel object behoort
- **Identificatie**: Het geheel aan cijfers, labels of codes die een object uniek identificeren
- **Data**: (Meet)gegevens van het object, of primaire gegevens die aan het object toebehoren
- **Metadata**: Eigenschappen het object, of secundaire gegevens die van belang zijn voor de context, kwaliteit, interpretatie of administratie van de primaire gegevens

2.2 Objecttypen of HydroObjects van de HydroMonitor

Als onderdeel van de investeringen en verbeteringen die KWR doorvoert in de programma's Menyanthes en HyCA voor respectievelijk waterkwantiteit en waterkwaliteit, is niet alleen het HydroMonitor data-uitwisselformat in het leven geroepen maar ook een nieuw, onderliggend datamodel. Uitgangspunt van beiden is dat er verschillende typen objecten of 'HydroObjects' worden onderscheiden die relevant zijn voor hydrologische analyses, modellering of monitoring. Omdat de verschillende HydroObjects in werkelijkheid elk hun eigen karakteristieken hebben, verschilt ook hun onderliggende, digitale weergave of datamodel. De HydroObjects die binnen de HydroMonitor worden onderscheiden zijn:

Meet- en boorlocaties:

1. Observation Well
2. Surface Water Level Gauge
3. Weather Station
4. Pumping Well
5. Borehole log

Meettechnieken:

6. Manual Measurement
7. Logger Measurement

Meetinstrumenten:

8. Logger
9. USB baro sensor

Gebeurtenissen:

10. Event

3 Beschikbare varianten

3.1 Spreadsheet- of matrixvariant (CSV en Excel, niet-hiërarchisch)

Voor de opzet van een data format is het van belang of de manier van dataopslag een hiërarchie toestaat, of niet. In het geval van tekst- of CSV- en Excel-bestanden is er geen (of slechts beperkt) hiërarchie mogelijk. Voor deze formats geldt dat een gegevenshiërarchie die van nature aanwezig en van belang is 'platgeslagen' moet worden in een 2D matrix of spreadsheetvorm. Dit resulteert in de kern in een opzet met verschillende tabellen die via de identificatie van objecten onderlinge relaties kunnen hebben waarmee de eventuele hiërarchie beschreven en vastgelegd kan worden. Op het hoogste niveau bestaan de HydroMonitor – CSV en Excel formats daarbij uit de volgende onderdelen (Figuur 1):

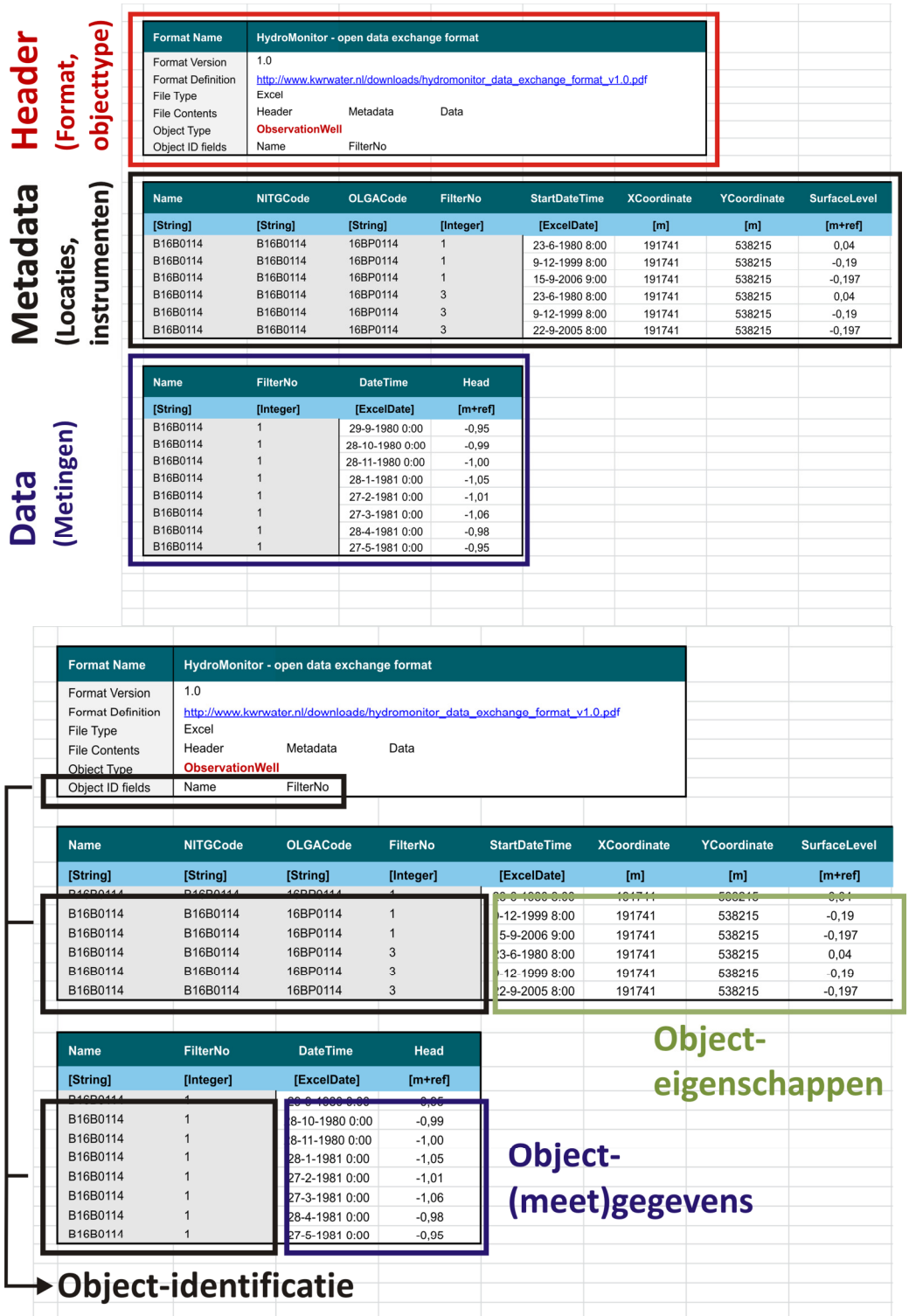
- **Header**: Een tabel met daarin de identificatie en eigenschappen van het bestand en format
- **Metadata**: Secundaire gegevens, of eigenschappen van bijvoorbeeld meetlocaties en meetinstrumenten
- **Data**: Primaire gegevens, bijvoorbeeld meetreeksen

In dit format kan een bestand gegevens van meerdere objecten bevatten, door deze onder elkaar op verschillende rijen of regels te zetten. Elke regel wordt daarbij voorzien van de unieke identificatie van het object, via de identificatie wordt bovendien een relatie gelegd tussen de verschillende elementen of tabellen uit het format. Voorbeelden van bestanden volgens het HydroMonitor CSV- en Excel-format zijn te vinden in de bijlagen A en B.

3.2 Binaire variant (MEF, hiërarchisch)

In verschillende binaire formats, waaronder MATLAB Exchange Format (MEF) bestanden die als basis het HDF5-format van MATLAB® gebruiken, is het wel mogelijk om de gegevens op te slaan in een hiërarchische vorm. In het geval van MATLAB gaat het daarbij bijvoorbeeld om zogenaamde 'structures', die als mogelijkheid hebben dat de ene structure genest wordt in de andere. Op het hoogste niveau heeft het HydroMonitor MEF-format de volgende opzet en hiërarchie (voor verdere toelichting en grafische weergave zie bijlage C):

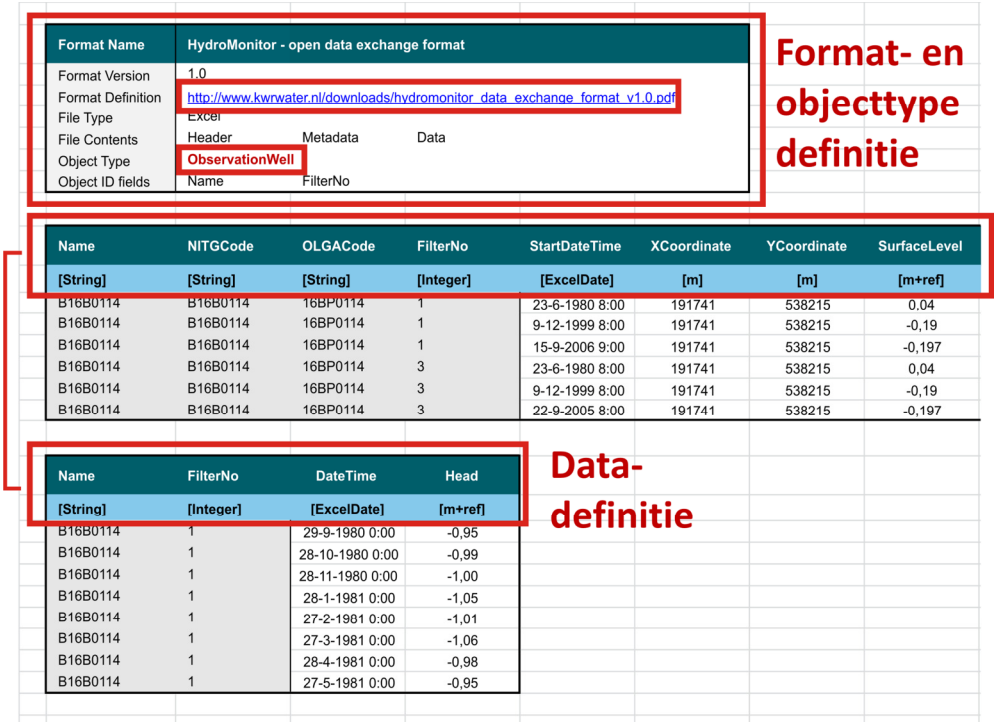
- **Database:** Een structure, die letterlijk de basis vormt waarop de overige gegevens rusten
  - **Format:** Een structure met daarin de identificatie en eigenschappen van het bestand en format
    - **Version:** Nummer dat de formatversie identificeert
    - **Definition:** Verwijzing naar een document met uitgebreidere format- en datadefinities
    - **Type:** Formattype
    - **Contents:** Inhoud van het bestand
  - **Object:** Een structure array met gegevens van de individuele objecten
    - **Identification:** Sleutel tot een individueel object
    - **Metadata:** Secundaire gegevens
    - **Data:** Primaire (meet)gegevens



Figuur 1: Toelichting op de opzet en verschillende onderdelen van het HydroMonitor Excel format (boven), en de manier waarop relaties worden gelegd via de identificatie van het object.

4 Gegevensdefinities

4.1 Format, datatype, eenheid en/of referentie



Figuur 2: In bestanden met het HydroMonitor format, Excel of anderszins, zijn de belangrijkste eigenschappen van bestand, format en aanwezige velden en kolommen (datatype en/of eenheid en referentie) opgenomen.

Fouten en misverstanden over de exacte definitie van gegevens kunnen verstrekkende gevolgen hebben (zie bijv. 'Some Famous Unit Conversion Errors', beschikbaar op http://spacemath.gsfc.nasa.gov/weekly/6Page53.pdf , of Von Asmuth en Van Geer, 2015). Om de kans op misverstanden te minimaliseren en het gebruik van het format te vergemakkelijken zijn in bestanden met het HydroMonitor format de belangrijkste eigenschappen van het bestand, format en de aanwezige velden en kolommen opgenomen (datatype en/of eenheid en referentie, zie Figuur 2). Voor meer gedetailleerde informatie is in het bestand bovendien een verwijzing of link opgenomen naar het onderhavige document met nadere toelichting en definities op internet.

Ondersteunde eenheden:
In de ideale situatie zou de software die de bestanden inleest ook de gebruikte eenheden uitlezen, en zo nodig converteren. Dit staat op de wensenlijst, maar is op het moment van schrijven nog niet geïmplementeerd in de HydroMonitor. De rij in de tabel die de gebruikte eenheden en referenties definieert, geeft weer wat deze zouden moeten zijn. Deze rij dient vooralsnog dus niet gewijzigd te worden!

Ondersteunde referentieniveaus en coördinatenstelsels:
Bij het invoeren van gegevens in Menyanthes en de HydroMonitor staat het u als gebruiker in principe vrij om zelf een referentieniveau en coördinatenstelsel te kiezen, zolang dat cartesisch is. Voor gebruik van de kaartinterface dienen het coördinatenstelsel van de meetlocaties en achtergrondtopografie overeen te komen. Voor het downloaden van luchtfoto's worden de coördinaten naar WGS84 getransformeerd. Op het moment van schrijven wordt deze transformatie alleen ondersteund voor Rijksdriehoeks(RD)- of Amersfoortcoördinaten, die binnen Nederland gangbaar zijn.

Ondersteunde datum- en tijdformats:
Het format waarmee datums en tijdstippen ingevoerd dienen te worden hangt af van het type bestand dat gebruikt wordt.
- Voor Excel bestanden is dit [ExcelDate], ofwel het interne datumnummer van Excel. Controleer zo nodig of een datum in Excel een datumnummer is door bijv. 1 dag erbij op te tellen.
- Voor CSV bestanden is dit [dd-mm-yyyy HH:MM]
- Voor MEF bestanden is dit [MatlabDate], ofwel het interne datumnummer van MATLAB.

4.2 Betekenis van velden en kolommen

Verschillende objecten zoals meetlocaties en meetinstrumenten hebben verschillende eigenschappen, waarvoor in databestanden verschillende velden of kolommen gereserveerd zijn. De paragraaf 'objecttypen of HydroObjects van de HydroMonitor' geeft een overzicht van de typen objecten die binnen de HydroMonitor onderscheiden worden. Voor elk van deze objecttypen zijn voorbeeldbestanden beschikbaar, waarin ook de velden en kolommen terug te vinden zijn die voor dat objecttype standaard beschikbaar zijn. We lichten de betekenis van die velden hieronder kort toe.

Velden van (meet)locaties in het algemeen:

Veld- of kolomnaam	Eenheid en referentie	Omschrijving
StartDateTime	[zie toelichting]	Ingangsdatum en tijdstip van de gegevens
XCoordinate	[m]	X-Coördinaat
YCoordinate	[m]	Y-Coördinaat
SurfaceLevel	[m+ref]	Maaiveldhoogte
VegTypo	[Categorical]	Vegetatietype
VegType	[String]	Vegetatietypologie
LoggerSerial	[String]	Serienummer van de datalogger
LoggerDepth	[-m+welltop]	Diepte van de sensor t.o.v. meetpunt
LoggerBrand	[Categorical]	Merk van de datalogger
LoggerType	[Categorical]	Type datalogger ('Unknown', 'Absolute' of 'Relative')
Comment	[String]	Opmerking
CommentBy	[String]	Persoon die de opmerking plaatste

Aanvullende velden voor peilbuizen en pompputten:

Veld- of kolomnaam	Eenheid en referentie	Omschrijving
WellTopLevel	[m+ref]	Bovenkant van de buis of meetpunt
FilterTopLevel	[m+ref]	Bovenkant van het filter
FilterBottomLevel	[m+ref]	Onderkant van het filter
WellBottomLevel	[m+ref]	Onderkant van de buis

Aanvullende velden voor peilschalen:

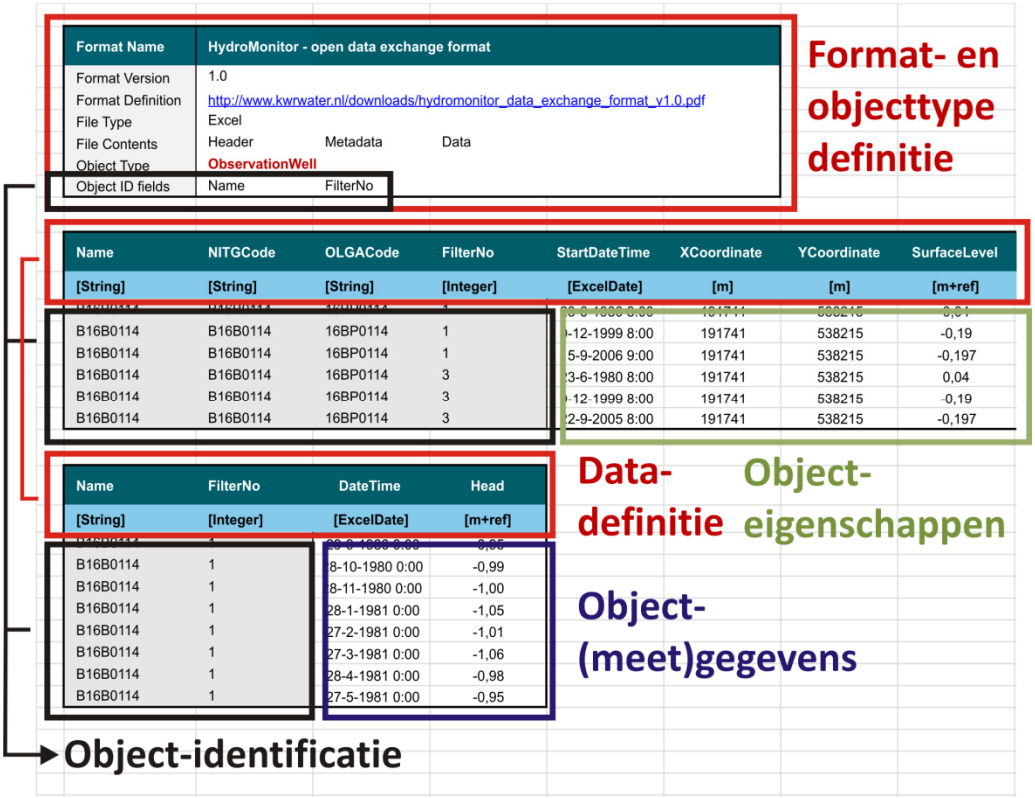
Veld- of kolomnaam	Eenheid en referentie	Omschrijving
GaugeTopLevel	[m+ref]	Bovenkant van de peilschaal of meetpunt
GaugeBottomLevel	[m+ref]	Onderkant van de peilschaal

Velden van meetreeksen:

Veld- of kolomnaam	Eenheid en referentie	Omschrijving
DateTime	[zie toelichting]	(Eind)datum en tijd van meting of flux
LoggerHead	[m+ref]	Grondwaterstand of stijghoogte, gemeten met logger
ManualHead, of Head	[m+ref]	Grondwaterstand of stijghoogte, gemeten met de hand
Level	[m+ref]	Waterstand
Volume	[m^3]	Watervolume, flux over de tijdstap
Precipitation	[mm]	Waterkolom, flux over de tijdstap
Evaporation	[mm]	Waterkolom, flux over de tijdstap
AirPressure	[hPa]	Luchtdruk
Temperature	[°C]	Temperatuur

Bijlage A: Het HydroMonitor Excel format

Vanwege de overzichtelijkheid en het gemak van Excel is het Excel format het meest geschikt voor gegevens die handmatig ingevoerd en/of gemanipuleerd moeten worden. Voor elk van de objecttypen van de HydroMonitor zijn voorbeeldbestanden beschikbaar, die meegeleverd worden bij installatie van de programmatuur. We tonen hieronder als voorbeeld de Observation Well variant, met illustratie van de verschillende onderdelen van het format zoals besproken in dit document.



Figuur 3: Voorbeeld van het HydroMonitor – Observation Well Excel format, met toelichting op de opzet en verschillende onderdelen daarvan.

Bijlage B: Het HydroMonitor CSV format

Het CSV format is gelijk aan het Excel format, met uitzondering van de opmaak, en kan vanuit en door Excel ook gemaakt of gelezen worden. Het is meest geschikte format voor uitwisseling van gegevens tussen verschillende programma's. Voor elk van de objecttypen van de HydroMonitor zijn voorbeeldbestanden beschikbaar, die meegeleverd worden bij installatie van de programmatuur. We tonen hieronder als voorbeeld een onderdeel van de Observation Well variant.

```
Format;HydroMonitor - open data exchange file;;;;;;;;;;;;;;
Format Version;1.0;;;;;;;;;;;;;;
Format Definition; www.menyanthes.nl/hydromonitor_data_exchange_format.pdf;;;;;;;;;;;;;;
Format Type;Excel;;;;;;;;;;;;;;
Format Contents;Header;Metadata;Data;;;;;;;;;;;;
Object Type;ObservationWell;;;;;;;;;;;;;;
Object ID Fields;NITGCode;FilterNo;;;;;;;;;;;;;
;;;;;;;;;;;;;;
Name;NITGCode;OLGACode;FilterNo;StartDateTime;XCoordinate;YCoordinate;SurfaceLevel;WellTopL
evel;FilterTopLevel;FilterBottomLevel;WellBottomLevel;LoggerSerial;LoggerDepth;Comment;Comm
entBy
[String];[String];[String];[Integer];[dd-mm-yyyy
HH:MM];[m];[m];[m+ref];[m+ref];[m+ref];[m+ref];[m+ref];[String];[-
m+welltop];[String];[String]
LocalName1;B52C0089;;1;15-6-1970 0:00;187930;378630;30,37;30,76;21;20;20;;;;
LocalName2;B52C0151;;1;15-6-1970 0:00;189790;382110;30,16;30,74;23,74;22,74;22,74;;;;
LocalName3;B52C0156;;1;15-6-1970 0:00;187360;377180;27,86;28,09;11,09;10,09;10,09;;;;
LocalName4;B52C0281;;1;14-11-1992 0:00;189440;383320;30,78;30,74;30,24;29,24;29,24;;;;
LocalName4;B52C0281;;2;1-6-2003      0:00;189440;383320;30,78;30,91;28,21;27,21;27,21;;;Buis
opgelengd;Jos von Asmuth
LocalName4;B52C0281;;2;14-11-1992 0:00;189440;383320;30,78;30,71;28,21;27,21;27,21;;;;
;;;;;;;;;;;;;;
NITGCode;FilterNo;DateTime;ManualHead;;;;;;;;;;;;
[String];[Integer];[dd-mm-yyyy HH:MM];[m+ref];;;;;;;;;;;;
B52C0089;1;27-5-1966 1:00;28,19;;;;;;;;;;;;
B52C0089;1;14-6-1966 2:30;28;;;;;;;;;;;;
B52C0089;1;28-6-1966 1:00;28,5;;;;;;;;;;;;
B52C0089;1;28-7-1966 1:00;28,45;;;;;;;;;;;;
B52C0089;1;15-8-1966 1:00;28,31;;;;;;;;;;;;
B52C0089;1;29-8-1966 1:00;28,19;;;;;;;;;;;;
B52C0089;1;14-9-1966 1:00;28,12;;;;;;;;;;;;
B52C0089;1;28-9-1966 1:00;28,03;;;;;;;;;;;;
B52C0089;1;14-10-1966 1:00;28;;;;;;;;;;;;
B52C0089;1;28-10-1966 1:00;28,34;;;;;;;;;;;;
B52C0089;1;14-11-1966 1:00;28,57;;;;;;;;;;;;
```

Bijlage C: Het HydroMonitor MEF format

De basis van het MEF format ligt in het binaire HDF5 format dat MATLAB gebruikt om gegevens in op te slaan en mee uit te wisselen (met als standaard extensie *.MAT en versie '-v7.3', zie ook http://nl.mathworks.com/help/matlab/import_export/mat-file-versions.html). Het MATLAB format is ook het format dat de HydroMonitor gebruikt als lokale database of databestand voor het integraal opslaan van alle gegevens (al dan niet als cache), maar de structuur van de eigen HydroMonitor bestanden is te zeer aan verdere verbetering onderhevig om als uitwisselformat te kunnen dienen. De interne opzet van het MEF format is vergelijkbaar met dat van de andere formats, in de zin dat er een header is met formatdefinities en ruimte voor de identificatie, metadata en data van de objecten. Het verschilt daarvan doordat het een hiërarchisch format is, waardoor meerdere objecten tegelijk gegroepeerd in een kolom opgeslagen kunnen worden. We tonen hieronder een grafische uitwerking van de structuur van dit format.

Eenheden en referentieniveaus in het MEF format:
In het MEF format bevinden de gegevens van objecten zich in de MATLAB-tabellen ID, Meta en Data. De eenheden en referentieniveaus van gegevens worden hier opgeslagen onder Properties.VariableUnits. Voorbeeld:

```
>> Meta.Properties.VariableNames
ans =
  Columns 1 through 5
'StartDateTime'    'XCoordinate'    'YCoordinate'    'SurfaceLevel'    'WellTopLevel'

>> Meta.Properties.VariableUnits
ans =
  Columns 1 through 5
    '[MatlabDate]'    '[m]'    '[m]'    '[m+ref]'    '[m+ref]'
```

Blijvende ondersteuning van MEF format versie 1.0:
We voorzien dat het MEF format in de (nabije) toekomst verder verbeterd en gewijzigd zal worden. Om data-uitwisseling via dit format tussen verschillende programma's en/of MATLAB-gebruikers te vergemakkelijken, zullen Menyanthes en de HydroMonitor in ieder geval versie 1.0 van dit format blijvend ondersteunen bij het importeren en exporteren van gegevens.

HydroMonitor – MEF example file v1.0.mef: MATLAB® HDF5 compliant binary file - MAT-file version '-v7.3' or compressed to optimize size

D: Structure (main database)

Header

Format: Structure (format definition)

Name	'HydroMonitor - open data exchange format'
Version	'1.0'
Definition	'http:/www.hydromonitor.nl/hydroMonitor_data_exchange_format.pdf'

File: Structure (file properties)

Type	'Matlab'
Contents	'Header', 'Objects'

Obj: Structure array [n x 1] (object data)

Type	'ObservationWell'

ID: Structure [1 x n] (identification)

Name	NITGCode	OLGACode	FilterNo
[String]	[String]	[String]	[Integer]
LocalName1	B52C0089		1

Meta: Structure [m x n] (properties)

StartDateTime	XCoordinate	YCoordinate	SurfaceLevel	WellTopLevel
[ExcelDate]	[m]	[m]	[m+ref]	[m+ref]
15-6-1970 0:00	187930	378630	30,37	30,76
15-6-1970 0:00	189790	382110	30,16	30,74
15-6-1970 0:00	187360	377180	27,86	28,09

Data: Structure [m x n] (measurements)

DateTime	Head
[ExcelDate]	[m+ref]
27-5-1966 1:00	28,19
14-6-1966 2:30	28
28-6-1966 1:00	28,5
28-7-1966 1:00	28,45
15-8-1966 1:00	28,31
29-8-1966 1:00	28,19