

# Data delivery for Marine Monitoring Projects

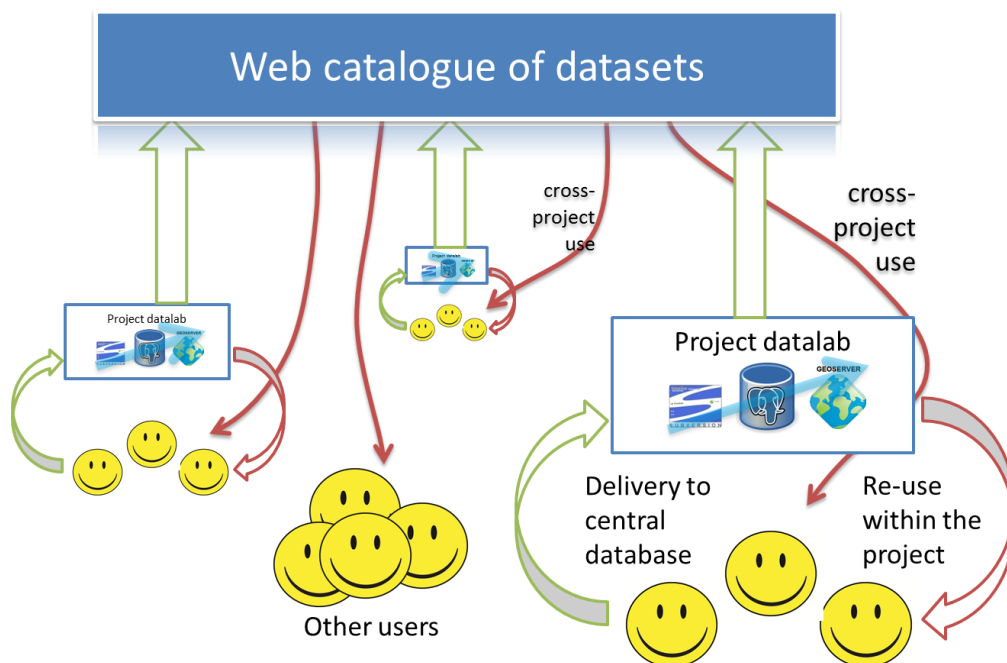
## Guidelines for data providers

Deltares, Wageningen Marine Research, Rijkswaterstaat

11 July, 2016

### 1 Background data management for Marine Monitoring Projects

Proper management of data within marine monitoring projects is essential for securing and efficient use of the data and information gathered in these large projects.



*Figure 1.1 - Proper data management will benefit the re-use of data within projects during their lifetime, and between projects, also after their lifetime.*

Efficient data management not only provides a means of optimal data and information exchange within the duration of the monitoring contract, but also facilitates the accessibility and reuse of the data after completion of the project.

This document provides guidelines for the delivery of data. It is a living document that can be adapted to include the development of standards used, changing needs, new insights, and comments from relevant parties. When in doubt about the use of these guidelines contact the data controllers of your project.

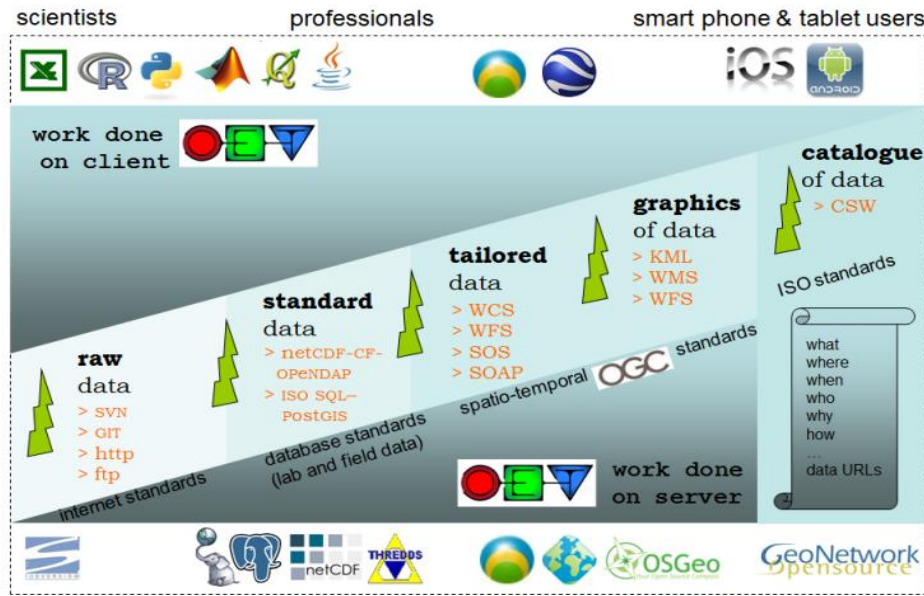


Figure 1.2 - Data management according to OpenEarth. For more information go to <http://www.openearth.eu/data>

The data management is performed according to the OpenEarth method and principles (<https://publicwiki.deltares.nl/display/OET/Data>). The data in the data management system are ultimately shown in a catalog as raw data, processed (standardized) data, or as a web service. This document is based on the generic "Data Protocol" (<https://publicwiki.deltares.nl/display/OET/Data+Protocol>) developed within OpenEarth.

Not all data collected in the monitoring projects will be centrally stored and managed. In consultation with RWS/WVL it will be indicated which data sets should be delivered.

The data provider is responsible for the supply of raw data (as is) and of base data (in a pre-defined data structure and format). Any reprocessing steps from the raw data that led to the delivered base data must be documented properly, preferably by delivery of scripts or mapping table.

### 1.1 Metadata

Metadata describing the data must be delivered in accordance with ISO 19155 and/or INSPIRE. It is recommended that you use the [INSPIRE metadata editor] (<http://inspire-geoportal.ec.europa.eu/editor/>). The metadata is put in the repository in a separate folder (see below). A template will be made available by the project data controller.

### 1.2 Feature data (points, lines, polygons, time series, profiles)

This data is expected two deliveries: *Raw data* and *Base data*

- **Raw data** is data originating from instruments or recorded observations from observers without any further processing. They may be delivered in any format that can be read by common programs (eg. CSV, MS Excel, MS Access, ESRI shape) or

together with the software to read the raw data. These data act as a back-up, and will not be further processed.

- **Base Data** is data that is the result of a first processing step towards a data product (e.g. aggregate, anonymize). Base data are more suitable for re-use by third parties. Base data are ingested in the database and further processed to deliver web services. Features of base data are:
  - Lowest resolution data (no data product)
  - Standardized according to "IMWA metingen"
  - Complete, to make all the products used in reporting
  - When in doubt, seek consultation with data controller (RWS and/or Deltares/WMR)

Base data should be delivered as one or more files (CSV, GML, JSON) formatted according to IM metingen model. See appendix for the precise size and an explanation of the different fields. This format is used in accordance with the "comply or explain" principle. In case information does not fit well in the format, it should still be added, with a detailed description. It is not intended to only deliver the required fields, if it leads to loss of information.

### 1.3 Grids or raster data (arrays)

Grids, raster data or arrays are accepted as NetCDF with complete metadata. It is recommended to use the [NetCDF kickstarter] (<http://zandmotor.citg.tudelft.nl/netcdfkickstarter/>) ([documentation] (<https://publicwiki.deltares.nl/display/OET/netCDF+kickstarter>)). The NetCDF kickstarter generates a script (Matlab, Python or R) for the production of a blank NetCDF file with metadata and the correct dimensions. The data must then be added in the script.

### 1.4 Uploading of data files

Data files are uploaded to the project's SVN repository. The repository of the web address is depending on the project. [This link](#) contains a tutorial for working with repositories. Contact your data controller for further information. The general structure within a repository looks like this:

- Project
  - Lot (e.g. fish, birds)
    - inspire\_metadata\_xml
    - raw\_data
    - base\_data (for standardized data )
    - scripts (e.g conversion of raw to base data)
    - documentation (e.g. reports, methodologies)

## 1.5 Support

Deltares/WMR provides support for the delivery of data, including converting data into the AQUO standard and uploading to the repository. For specific questions about AQUO standards we refer to Information Water House, the administrator of aquo.

## 1.6 References and links

Information about aquo exchange formats in general: <http://www.aquo.nl/over-aquo/aquo-onderdelen/aquo-uitwisselformaten/>

Source IM metingen: <http://www.aquo.nl/over-aquo/aquo-onderdelen/aquo-modellen/imwa-metingen/>

Examples of IM measurements csv files: <http://www.aquo.nl/over-aquo/aquo-onderdelen/aquo-uitwisselformaten/csv-formaat/#csv-encoding-imwa-metingen-en-waterveiligheid>

AQUO code list vocabularies: <http://domeintabellen-idsw.rws.nl>

Practical information about AQUO code lists

<http://www.aquo.nl/documents/2016/04/aquo-praktijkrichtlijn-domeintabellen-imwa-metingen.pdf>

NetCDF kickstarter: <http://zandmotor.citg.tudelft.nl/netcdfkickstarter/>

NetCDF kickstarter documentation:

<https://publicwiki.deltares.nl/display/OET/netCDF+kickstarter>

Explain work with SVN repositories:

[https://publicwiki.deltares.nl/download/attachments/42401943/how\\_to\\_use\\_subversion.pps?version=5&modificationDate=1320923358000&api=v2](https://publicwiki.deltares.nl/download/attachments/42401943/how_to_use_subversion.pps?version=5&modificationDate=1320923358000&api=v2)

INSPIRE metadata editor: <http://inspire-geoportal.ec.europa.eu/editor/>

## 1.7 Appendix

Table 1. IM fields, examples and further explanation. Further documentation in <http://www.aquo.nl/documents/2015/04/rapport-imwa-metingen-2015.pdf>

veldnaam	Voorbeeld	RWS	opmerking / toelichting
		vulling van het veld is: v=verplicht, c=conditoneel, o=optioneel	
Meetpunt.identificatie	NOORDWK10	v	Id nummer van meetobject (of locatie of trek) waar "monster" en "meting" naar verwijzen
Metingomschrijving	Noordwijk 10 km uit de kust	o	naam van meetobject, locatie of trek bijv Noordwijk-10
GeometriePunt.x	gg.ggggg	c	de x en y in etrs89. (wgs84 mag ook als benadering gebruikt worden, omdat de afwijking op de Noordzee max 80 cm is) Bij een trek of een raai kan het middelpunt aangegeven worden of een polygoon. Indien polygoon gewenst zijn, kan dat met een string worden aangegeven in het veld "geometrie",
GeometriePunt.y	gg.ggggg	c	
Geometrie		c	(vlak, lijn, raster, polygoon) als de meetlocatie geen punt is
Referentiehorizontaal.code	EPSG4258 of EPSG:4258	v	EPSG4258 (=etrs89) Op dit moment accepteert aquo geen : maar dat is internationaal wel gebruikelijk. Hiervoor is een change verzoek ingediend bij AQUO in dec door Joan Staeb
Monster.identificatie	Imaresnr_2014_43256421	v	unieke identificatie monster (max 36 tekens)
Compartiment.code	BS	v	~28 domeinwaarden zoals: oppervlaktewater, organisme (biota) etc
Orgaan.code	FI	c	indien organisme dan verplicht ~32 domeinwaarden

Organisme.naam	Macoma balthica	c	zoals: spierweefsel, dielijk weefsel etc indien organisme dan verplicht ~duizenden domeinwaarden zoals Abra, Abra alba, etc.
Begindiepte_m	-1	c	monsternemingsdiepte t.o.v. referentievlak (-1 betekent 1m onder referentievlak; +1 betekent 1 m boven referentievlak)
Einddiepte_m	-1	c	verplicht als een begindiepte is opgegeven.
Referentievlak.code	WATSGL	c	~28 domeinwaarden zoals: tov mean sea level, tov NAP, tov bodem, tov WATSGL
Monsterbewerkingsmethode.code	ISO5667-3	o	~23 domeinwaarden zoals: NEN-EN-ISO 5667-3, etc
Bemonsteringsmethode.code	rws huismethode 231	v	~53 domeinwaarden zoals: NENs en RWSVs
Bemonsteringsmethode.codespace	www.rws.nl	c	Indien hierboven een vrije waarde is ingevuld (dus geen domeinwaarde) dan moet de namespace worden ingevuld. Bijv www.rws.nl of www.WMR.nl
Monstercriterium.code		o	De domeintabel bestaat uit twee subtabellen van de domeintabel Hoedanigheid: BiologischeKenmerk en Korrelgroottefractie. om een nadere classificatie en/of beperking van het monster te geven. Veel beperkingen kunnen al via de opdeling van het monsterobject verder worden uitgewerkt, maar in sommige gevallen zoals bij biologische metingen is het noodzakelijk om aan te geven dat er enkel een specifiek onderdeel bemeten is. Bijv: alleen bepaalde vislengtes, of van bodemmonster wordt alleen het fijne deel geanalyseerd.
Veldapparaat.omschrijving	Boxcorer	v	De domeintabel Veldapparaat bestaat uit de subtabellen 'Meetapparaat', 'Bemonsteringsapparaat' en 'Plaatsbepalingsapparaat'. Veelal zal het gaan om een bemonsteringsapparaat. Er zijn ~67 domeinwaarden zoals: bodemschaaf, boxcorer, etc.
Monsternemingsdatum	2014-06-01	v	datum in jjjj-mm-dd (denk aan de verplichte nullen!!)
Monsternemingstijd	09:44	o	tijd als hh:mm of hh:mm:ss indien seconden relevant zijn
Tijd_UTCOffset	+1	c	Als monsternametijd is ingevuld en niet in MET zijn opgenomen (UTC / GMT + 1) dan hier de afwijking ten opzichte van UTC opnemen (wintertijd = MET, zomertijd =

			UTC + 2)
Typering.code		c	~57 domeinwaarden, subdeel van domeintabel parameter , je hebt of een typering, of een grootheid
Grootheid.code	AANTPOPVTE	c	je hebt of een typering, of een grootheid
Parameter.code		c	~3000 domeinwaarden (bij chemische stoffen en objecten verplicht)
Parameter.omschrijving	Macoma balthica	c	~3000 domeinwaarden, bij chemische stoffen is de code bepalend en is dit alleen een optionele toelichting: bij biotaxon is dit veld verplicht
Eenheid.code	mg/m2	v	~167 domeinwaarden
Hoedanigheid.code	adg	o	~415 domeinwaarden verdeeld in referentievlakken, fracties, equivalenten, referentiebases, biologische kenmerken en combinaties hiervan (bijvoorbeeld: asvrij drooggewicht: adg)
Waardebewerkingsmethode.code		o	~50 domeinwaarden verdeeld in percentielen en statistiek
Waardebepalingsmethode.code	Imaresmethode 2014-781	v	~600 domeinwaarden, vooral NENs en ISOs. Huismethodes dienen opgenomen te worden als "other: Naam van de huismethode incl nr en versie"
Waardebepalingsmethode.codespace	www.WMR.nl	c	Indien hierboven een vrije waarde is ingevuld (dus geen domeinwaarde) dan moet de namespace worden ingevuld. Bijv www.rws.nl of www.WMR.nl
Waardebepalingstechniek.code		o	~150 domeinwaarden, laboratoriumprincipes, geven globaal idee van de techniek. Van belang voor mapping naar internationale standaarden
Begindatum	2014-06-01	v	Datum waarover de waarneming geldig is. In de praktijk meestal de monsternemingsdatum
Begintijd	09:44	o	
Einddatum		o	Vooral nodig bij tijdsproportionele bemonsteringen.
Eindtijd		o	
Tijd_UTCOffset	+1	c	Als begintijd is ingevuld en niet in MET zijn opgenomen

Resultaatdatum	2014-06-01	v	(UTC / GMT + 1) dan hier de afwijking ten opzichte van UTC opnemen (wintertijd = MET, zomertijd = UTC + 2) Datum waarop de meting is gedaan (dit veld is nieuwe ISO O&M verplichting). Tijdelijke oplossing Indien dit niet bekend is wegens ontbreken in de aanleverende database dan laatste datum uit bovenstaande velden overnemen
Limietsymbool		c	"<" indien waarneming beneden de rapportagegrens is, ">" indien waarneming boven het bereik is.
Numeriekewaarde	24.2	c	De waarneming!!, Let op alleen significante decimalen meegeven
Alfanumeriekewaarde		c	indien geen numerieke waarde is gegeven
Kwaliteitsoordeel.code	0	v	~13 domeinwaarden. Uit mappingwerkgroep is gebleken dat alleen waarde 0 en 3 betekenis hebben.
Accreditatiestatus.omschrijving	ISO17025	o	ISO17025;geen;ander
Meetonzekerheid.bias.waarde	+10	o	
Meetonzekerheid.reproduceerbaarheid.waarde	30	o	2 sigma waarde
Meetonzekerheid.eenheid.code	%	o	%, ug/L, mg/L, pH etc
Meetonzekerheid.referentie.waarde	10	c	indien hierboven % dan hier referentiewaarde opgeven
Meetonzekerheid.referentie.eenheid	n	c	indien hierboven waarde ingevuld, dan hier de eenheid: ug/L, pH etc