

XHydro UML - Dokumentation

INHALTSVERZEICHNIS

1	XHYDRO UML - MODELL	1
2	ELEMENTE DES MODELLS.....	2
2.1	ZEITREIHENKERN	3
2.1.1	<i>Datentypen</i>	4
2.2	METADATEN	4
2.2.1	<i>Herkunft der Daten</i>	5
2.2.2	<i>Art der Daten</i>	6
2.2.2.1	Messgröße.....	6
2.2.2.2	Maßeinheiten	6
2.2.2.3	Bildungsregel	7
2.2.3	<i>Qualitätsangaben</i>	8
3	ABBILDUNG UML-BEZEICHNUNG ZU XML-SCHEMA-ELEMENTNAME	9
4	LITERATUR.....	13

Abbildungsverzeichnis

ABBILDUNG 1: UML - DIAGRAMM DES MODELLS	2
ABBILDUNG 2: MODELL DES ZEITREIHENKERNS	3
ABBILDUNG 3: DATENTYPEN.....	4
ABBILDUNG 4: DATENHERKUNFT.....	5
ABBILDUNG 5: DATENARTEN	6
ABBILDUNG 6: BILDUNGSREGEL	7
ABBILDUNG 7: QUALITÄTSANGABEN	8
ABBILDUNG 8: GEGENÜBERSTELLUNG UML-BEZEICHNUNG ZU XML-SCHEMA-ELEMENTNAME.....	9
ABBILDUNG 9: UML-DIAGRAMM MIT XML-ELEMENTNAMEN	12

1 XHydro UML - Modell

„XHydro“ ist ein XML - Schema [\[1\]](#), welches für die Belange der Datenübertragung im hydro-logischen Messnetz der Wasser- und Schifffahrtsverwaltung konzipiert wurde.

Im Folgenden wird die Struktur von XHydro (Version 1.3) anhand eines UML - Modells [\[2\]](#) dargestellt.

Bei der Entwicklung des XML - Schemas wurde zunächst ein generisches UML - Modell entworfen, von dem anschließend das hier beschriebene fachliche Konstrukt abgeleitet wurde. Basierend auf

diesem Modell wurde dann das XML - Schema XHydro generiert.^a

Durch den modularen Aufbau des XML - Schemas ist es leicht möglich, Anpassungen für andere fachliche Anforderungen, z.B. durch Austausch von Codelisten, vorzunehmen. Die einzelnen in diesem Modell enthaltenen Metadaten sind spezifisch für hydrologische Zeitreihen. Der Kern des Modells ist jedoch unabhängig von der Anwendungsdomäne und kann für verschiedene Typen von Zeitreihen wiederverwendet werden. Durch diese Flexibilität wird zudem erreicht, dass für die Datenübertragung im Messnetz ein weniger umfangreiches Schema verwendet werden kann als bei Datenarchivierungen oder beim Datenaustausch aus Datenbanken. Zudem wurde das XML - Schema, in Ergänzung zum UML - Modell so gestaltet, dass es sich fast beliebig um fachspezifische Elemente erweitern lässt. Das XML - Schema wird in [3] dokumentiert.

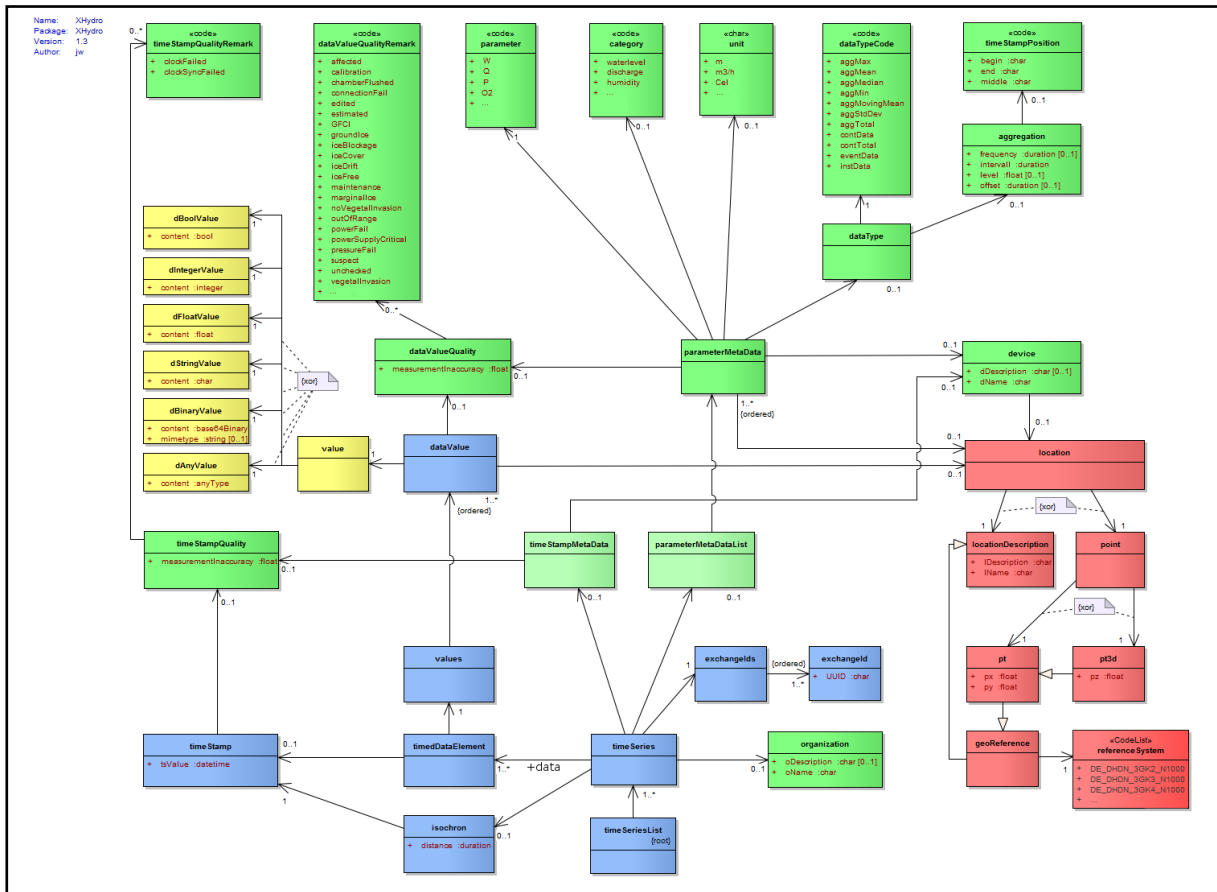


Abbildung 1: UML - Diagramm des Modells^b

2 Elemente des Modells

Die Elemente des Modells lassen sich 2 Hauptgruppen zuordnen:

- Zeitreihenkernel
- Metadaten

Der Zeitreihenkernel, in obiger Abbildung sind dies die blau dargestellten Elemente, bildet im Wesentlichen das grundsätzliche Konzept von Zeitreihen ab, d.h. die Zuordnung von Werten zu einem

^a Da die von UML-Werkzeugen automatisch generierten XML-Schemata die Anforderungen hinsichtlich Modularität und Wiederverwendbarkeit nicht erfüllten, erfolgte diese Umsetzung manuell.

^b Das UML - Modell enthält nicht die kompletten Codelisten, diese sind erst im XML - Schema vollständig. Auch die Erweiterbarkeit um Anwendungs- bzw. Nutzer- spezifische Elemente ist im UML - Entwurf noch nicht enthalten.

bestimmten Zeitpunkt.

Die rot und grün dargestellten Elemente sind weitergehende Merkmale. Diese Metadaten beschreiben die Zeitreihe und schaffen die Voraussetzung zur korrekten Interpretation der Daten bzw. deren Einordnung in einem Zeitreihen-Informationssystem.

2.1 Zeitreihenkernel

Das Modell erlaubt mehrere Zeitreihen in einer Liste zusammenzufassen. Jede Zeitreihe besitzt dabei im Kern jeweils die Elemente **timeSeriesList**, **timeSeries**, **exchangeIs**, **exchangeId**, **data**, **timedDataElement**, **timeStamp**, **values**, **dataValue** und ggf. **isochron**.

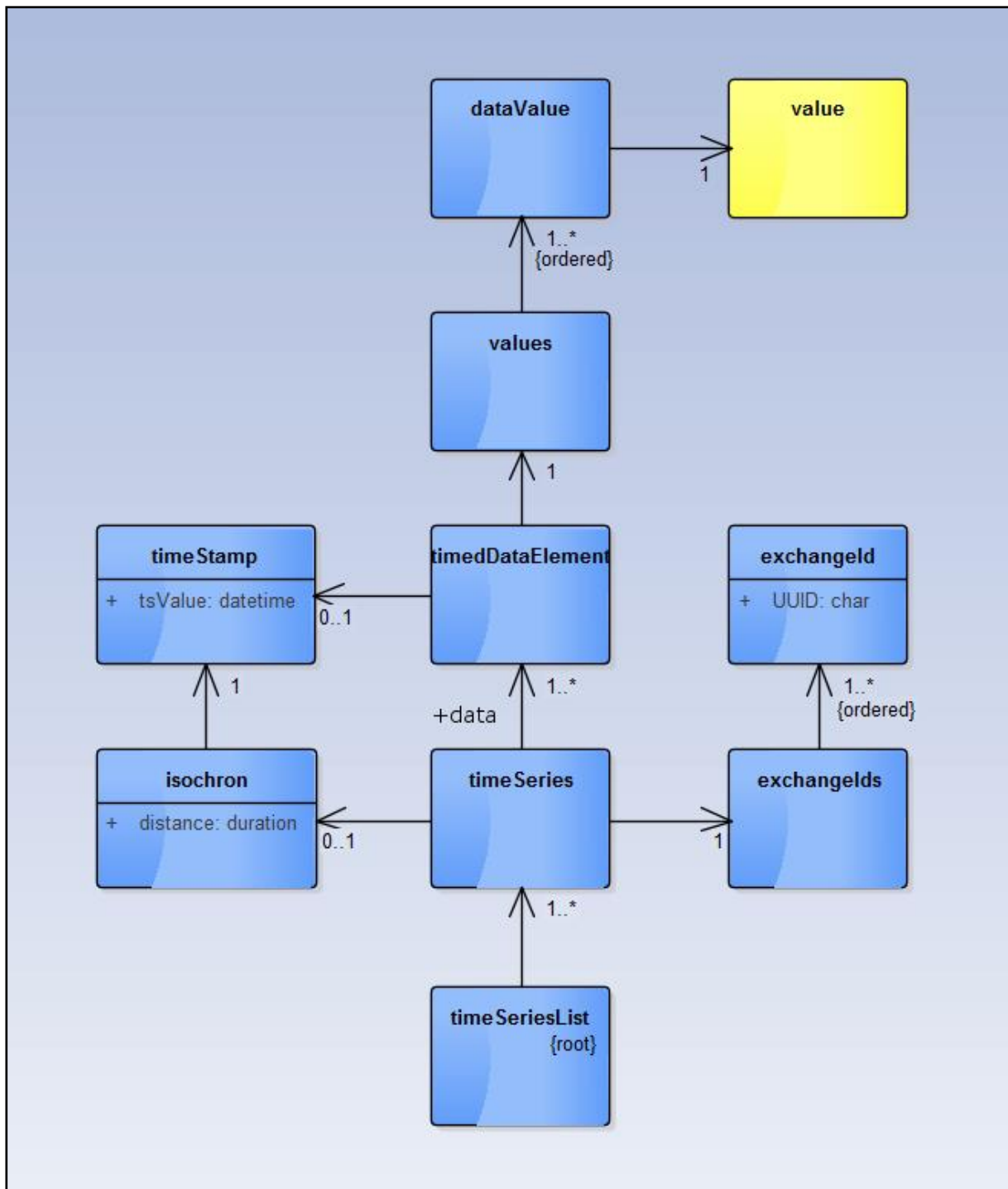


Abbildung 2: Modell des Zeitreihenkernelns

Das Element **timeSeriesList** ist das Root - Element und verweist auf ein oder mehrere Zeitreihenelemente. Das Zeitreihenelement **timeSeries** enthält ein **data**-Element mit einer Liste

aus einem oder mehreren **timedDataElement** - Objekten. Jedes **timedDataElement** - Objekt besteht aus einem Zeitstempel **timeStamp** - und einem **values** - Element.

Das **values** - Objekt kann ein oder, zum Zweck einer kompakten Kodierung, mehrere **dataValue** - Elemente beinhalten. Dies entspricht im logischen Aufbau einer Tabelle mit einer Zeitstempel- und einer oder mehreren Messwert-Spalten; jede Tabellenzeile entspricht hierbei einem **timedDataElement** - Objekt.

Das ggf. vorhandene Element **isochron** zeigt an, in welchem zeitlich äqui-distanten Abstand zueinander die Daten vorliegen, sowie zu welchem Zeitpunkt die Zeitreihe beginnt. Dies ermöglicht eine kompakte Darstellung, indem auf die Angabe expliziter Zeitstempel für jeden einzelnen Wert verzichtet werden kann. Der Zeitstempel jedes Wertes berechnet sich in solch einem Fall aus seiner Position in der Zeitreihe und dem Zeitinkrement (**duration**) sowie dem Startzeitpunkt (**startTimeStamp**).

Für den Austausch von Zeitreihen ist es sinnvoll, diese jeweils mit einem [eindeutigen](#) Identifikator zu kennzeichnen. Dies vereinfacht die Zuordnung der Daten und den Import in ein Zeitreihen-Informationssystem. In Version 1.3 des XML - Schemas werden hierzu im **exchangeID** - Element sog. Universally Unique Identifier [4] (UUID, hier nach [Version 4](#)) verwendet.

2.1.1 Datentypen

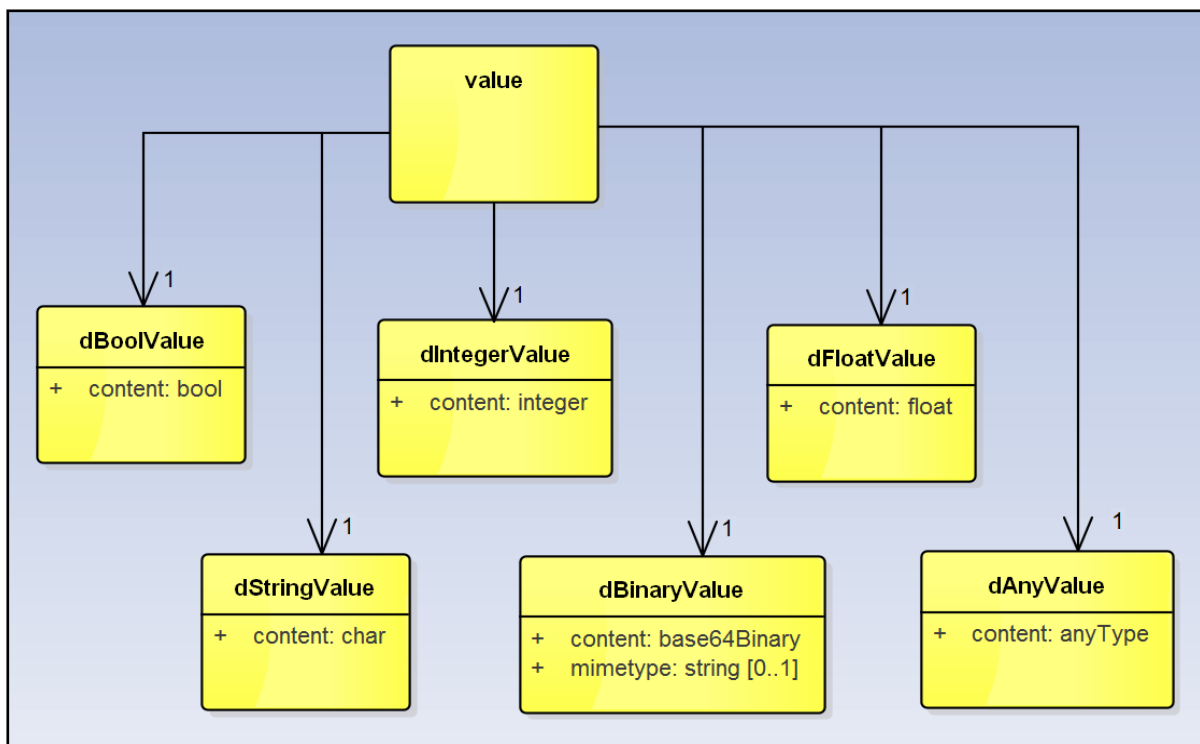


Abbildung 3: Datentypen

Daten hydrologischer Zeitreihen können von ganz unterschiedlichem Typ sein. Das Modell bietet die hierbei benötigten Datenformate an. Messwerte können als Fließ-komma- (**dFloatValue**) oder Ganzzahlen (**dIntegerValue**) dargestellt werden. Ja/Nein - Aussagen, wie Zustände an Anlageteilen, lassen sich mittels des logischen Typs (**dBoolValue**) und Logbucheinträge in Form von Zeichenketten (**dStringValue**) speichern. Komplexere binäre Datenobjekte sind mittels des Datentyps **dBinaryValue** kodierbar, wobei das Unterelement **mimetype** [5] spezifiziert, um welche Art von Datenobjekt (z.B. image/JPEG, video/mp4, application/pdf, audio/mpeg) es sich handelt.

2.2 Metadaten

Metadaten beschreiben die jeweilige Zeitreihe und liefern Informationen zu folgenden Kategorien:

- Herkunft der Daten
- Art der Daten
- Qualität der Daten

Das Modell benutzt hierzu u.a. mehrere Codelisten, welche jeweils aus einer vor-gegebenen Menge von Einträgen bestehen, die sich jederzeit erweitern lassen. Die konkrete technische Umsetzung des Erweiterungsmechanismus ist im XML - Modell beschrieben.

2.2.1 Herkunft der Daten

Im Modell kann die Herkunft der Daten sowohl über eine informelle Beschreibung des Messgeräts und/oder der Örtlichkeit erfolgen, als auch durch präzise Angabe der Messposition. Darüber hinaus kann beschrieben werden, von welcher Organisation / Institution die Daten stammen.

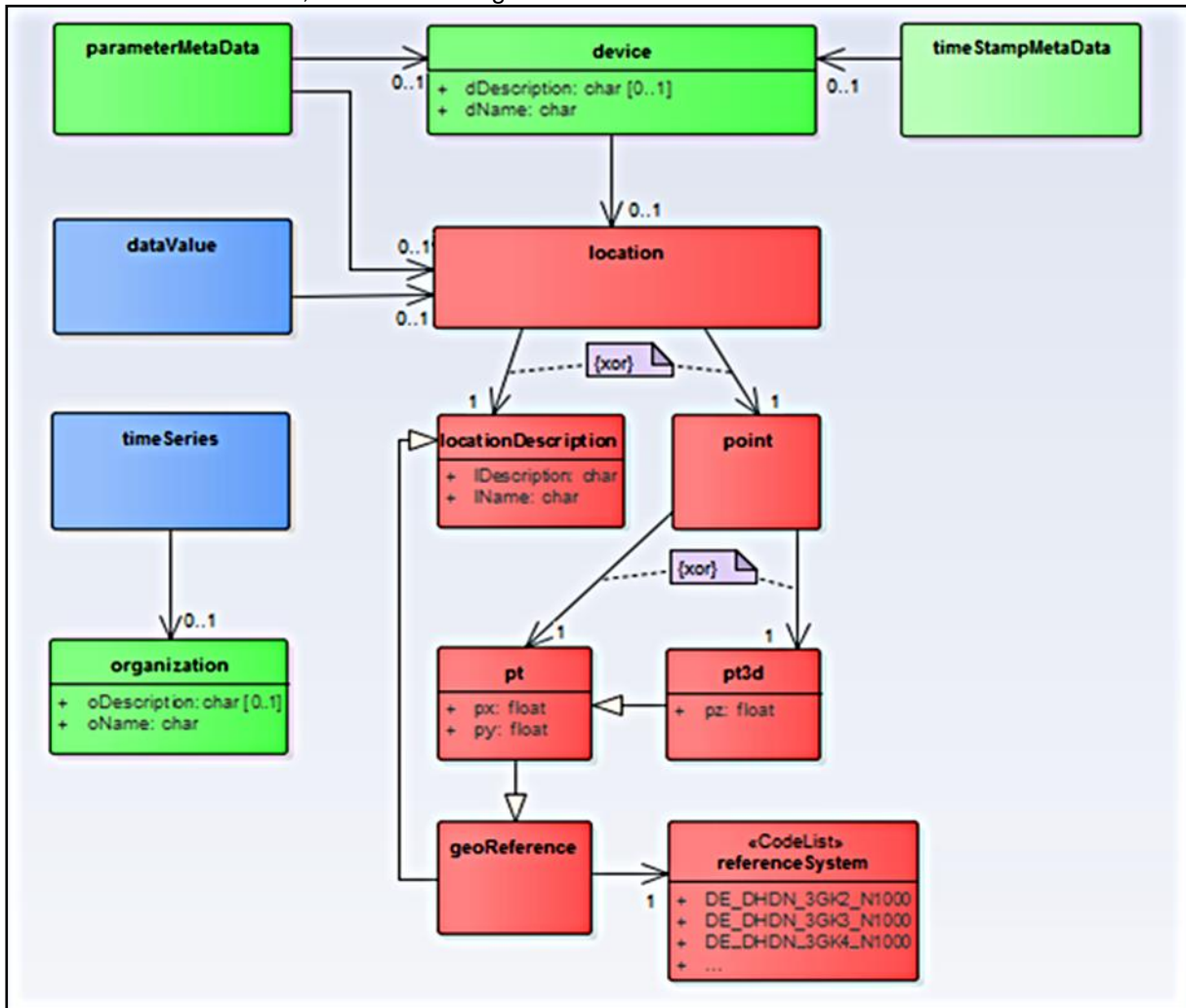


Abbildung 4: Datenherkunft

Das Element **device** kann hierbei zur Beschreibung des Messgerätes bzw. Sensors sowie der verwendeten Systemuhr verwendet werden. Angaben zum Messort^c als auch zum Installationsort des Sensors / der Systemuhr liefert **location**, wobei dieses Element entweder nur eine informelle Beschreibung in **locationDescription** beinhaltet oder diese um eine 2D- bzw. 3D- Ortsangabe incl. Nennung des verwendeten Referenzsystems ergänzt.^d

Diese Metadaten können individuell jedem einzelnen Messwert und/oder global der gesamten Zeitreihe zugeordnet werden.

Das Element **organization** beschreibt die Zugehörigkeit der Zeitreihen zur Organisation / Institution welche die Daten gemessen oder berechnet hat.

^c Der Messort ist die Örtlichkeit, für die die physikalische Größe bestimmt werden soll.

^d Zwar kann das **location** - Element nur entweder ein **locationDescription**- oder ein **point**- Element enthalten. Da aber **locationDescription** eine Basisklasse von **point** darstellt, ist die Möglichkeit zur informellen Ortsbeschreibung immer gegeben.

2.2.2 Art der Daten

Die Art der Daten wird beschrieben durch:

- Messgröße
- Maßeinheit
- Bildungsregel

2.2.2.1 Messgröße

Die Messgröße wird durch die drei Codelisten **parameter**, **category** und **unit** beschrieben.

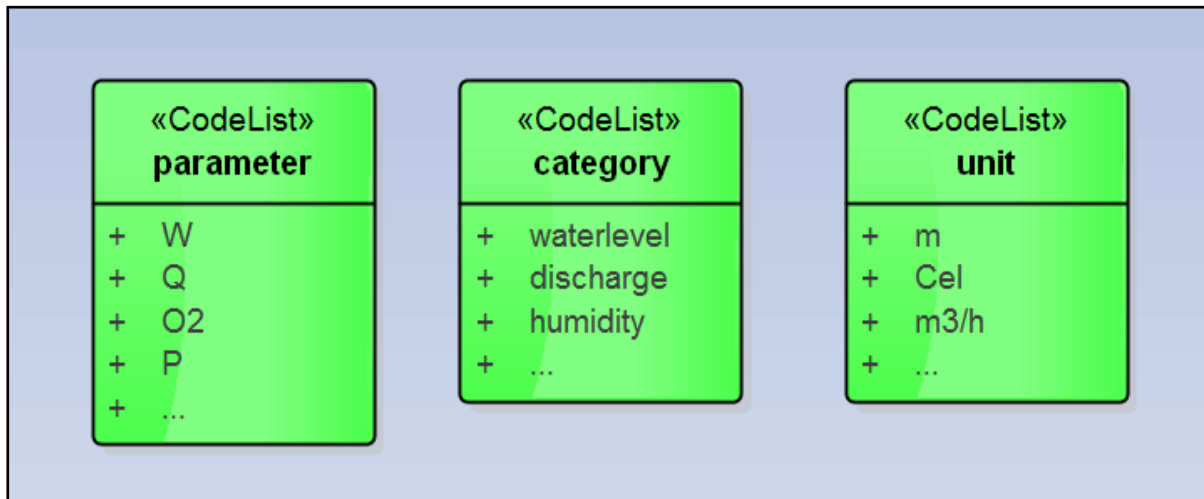


Abbildung 5: Datenarten

Das **parameter** - Objekt beschreibt, um welche Messgröße es sich handelt. Vordefinierte Parameter sind z.B. **W** für Wasserstand, **Q** für Abfluss und **T_W** für Wassertemperatur. Das **category** - ähnelt dem **parameter** - Objekt, hier wird aber nur die Hauptkategorie der Messgröße spezifiziert (z.B. **temperature** für den Parameter **W_T**).

2.2.2.2 Maßeinheiten

Das Attribut **unit** beschreibt die physikalische Maßeinheit der Messgröße. Die vordefinierte Codeliste orientiert sich am „Unified Code for Units of Measure“ - Dokument [6] und benutzt die „case sensitive“ - Schreibweise. Weitere, d.h. nicht vordefinierte Bezeichner können entsprechend dieser Nomenklatur verwendet werden.

2.2.2.3 Bildungsregel

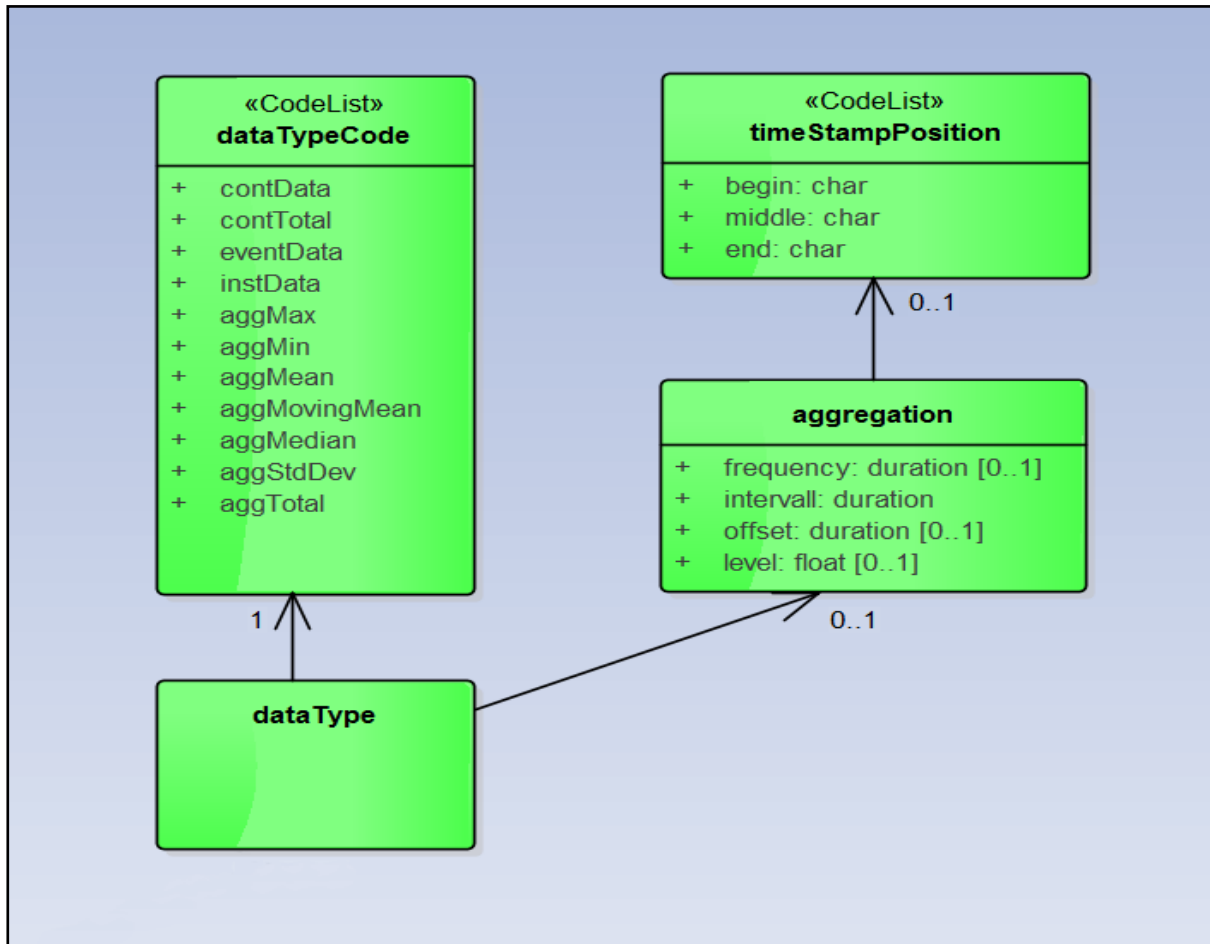


Abbildung 6: Bildungsregel

Durch diese Bildungsregel-bezogenen Codelisten wird beschrieben, wie der Wert entstanden ist.

Der Parameter **dataTypeCode** beschreibt die Bildungsregel des Messwertes:

contData	kontinuierlich (isochron oder zeitl. unregelmäßig) gemessen
contTotal	kontinuierlich integrierte oder hochgezählte Werte
eventData	ereignisgesteuerte Messung
instData	Instantanwert (d.h. keine Aggregation mehrerer Werte)
aggMax	höchster Messwert eines Beobachtungsintervalls
aggMin	niedrigster Messwert eines Beobachtungsintervalls
aggMean	Mittelwert aller Messwerte eines Beobachtungsintervalls
aggMovingMean	gleitender Mittelwert aller Messwerte eines Beobachtungsintervalls
aggMedian	Medianwert aller Messwerte eines Beobachtungsintervalls
aggStdDev	Standardabweichung aller Messwerte eines Beobachtungsintervalls

“**cont**” bedeutet, dass Messwerte kontinuierlich, isochron oder in zeitlich unregelmäßiger Folge, gemessen werden.

“**event**” bedeutet, dass Messwerte ereignisgesteuert, typischerweise bei Über- bzw. Unterschreitung fester Schwellwerte, oder bei Überschreitung einer kritischen Messwertveränderung gemessen werden.

“**agg**” bedeutet, dass über einen gegebenen Zeitraum mehrere Messungen erfolgen, woraus eine aggregierte Größe gebildet wird.

Nähere Informationen zum Aggregationsprozess liefert das Element **aggregation**, seine Attribute sind:

Frequency	beschreibt die Frequenz mit der (Zwischen-) Messwerte ermittelt werden; angegeben wird hierbei deren zeitlicher Abstand
Interval	gibt das Aggregationsintervall an, d.h. die Zeitspanne für die der Messwert repräsentativ ist
Offset	gibt an, mit welchem zeitlichen Offset (zur jeweils nächstgrößeren Zeiteinheit) das Messintervall beginnt; z.B. bedeutet ein Offset von 5min bei einer Intervalllänge von 30min, dass das Zeitintervall jeweils 5min und 35min nach jeder vollen Stunde beginnt
Level	beschreibt im Falle einer ereignisgesteuerten Zeitreihe den Schwell- oder Deltawert, dessen Über- bzw. Unterschreitung eine Messwert-speicherung triggert

Das Element **timeStampPosition** beschreibt die Position des Zeitstempels im Zeitintervall, d.h. ob der Zeitstempel den Beginn (**begin**), das Ende (**end**) oder die Mitte (**middle**) des Zeitintervalls markiert.

2.2.3 Qualitätsangaben

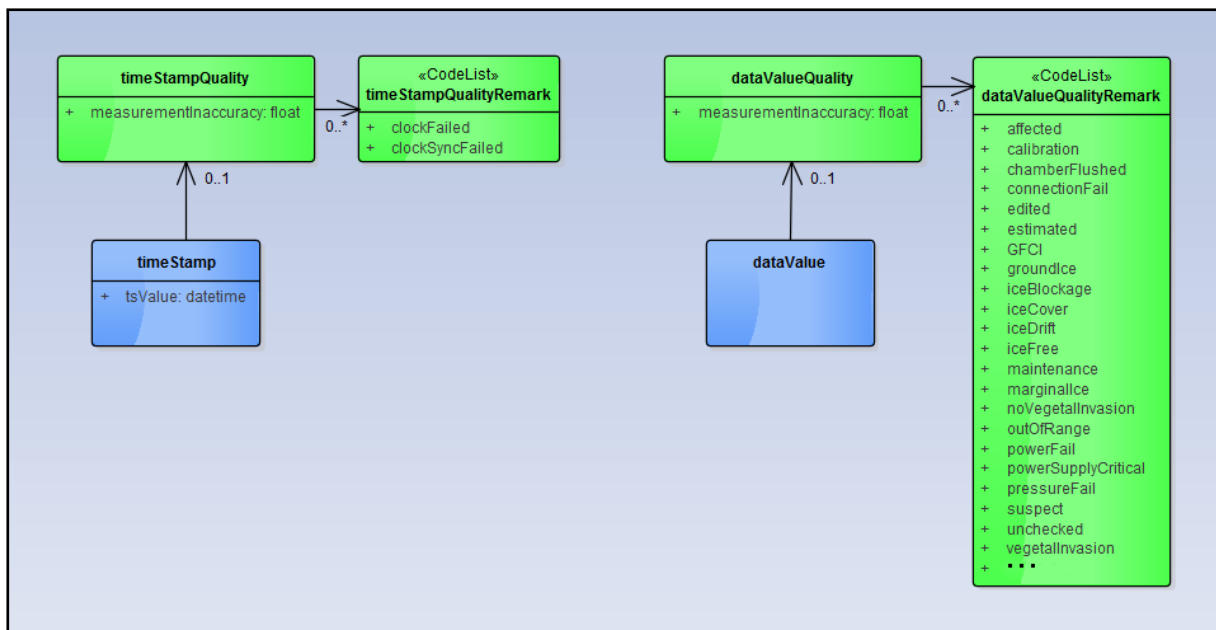


Abbildung 7: Qualitätsangaben

Das Modell sieht Qualitätsangaben sowohl für den Messwert als auch den Zeitstempel vor, wobei das Element **measurementInaccuracy** eine quantitative Angabe zur Messunsicherheit [7] enthalten kann. Die Qualität des Messwertes bzw. des Zeitstempels kann darüber hinaus auch durch einen Kommentar beschrieben werden. Die Codelisten **dataValueQualityRemark** bzw. **timeStampQuality** enthalten hierzu entsprechende Standardkommentare, wobei sich auch diese Codelisten jederzeit erweitern lassen.

3 Abbildung UML-Bezeichnung zu XML-Schema-Elementname

UML-Bezeichnung	XML-Schema-Element	Bedeutung
<i>(nicht in UML modelliert)</i>	ext	Erweiterungselement, bietet an vielen Stellen die Möglichkeit beliebige Zusatzinformationen hinzuzufügen
aggregation	ag	Informationen über den Aggregationsprozess
category	c	Hauptkategorie des Messwertes (z.B. „Temperatur“ für die parameter „Wassertemperatur“ bzw. „Lufttemperatur“ etc.) (Freitext)
category	xc	Hauptkategorie des Messwertes (z.B. „Temperatur“ für die parameter „Wassertemperatur“ bzw. „Lufttemperatur“ etc.) (Eintrag aus Codeliste)
content <i>(Kindelement von dBinaryValue)</i>	b	Binär kodierter Wert
dAnyValue	va	Selbst definierter Datentyp (kann beliebig komplex strukturiert sein)
data	d	Enthält alle timedDataElement / <tde> -Elemente der Zeitreihe
dataType	dt	Information ob und ggf. welche aggregierte Größe vorliegt
dataTypeCode	dtc	Informationen wie der Messwert gebildet wurde (Freitext)
dataTypeCode	xdtc	Informationen wie der Messwert gebildet wurde (Eintrag aus Codeliste)
dataValue	v	Enthält die Messgröße und die Qualitätsangaben
dataValueQuality	vq	Informationen zur Messunsicherheit und besonderen Mess-Umständen
dataValueQualityRemark	vqr	Informationen über besondere Mess-Umstände (Freitext)
dataValueQualityRemark	xvqr	Informationen über besondere Mess-Umstände (Eintrag aus Codeliste)
dBinaryValue	vbin	Binär kodierter Wert (z.B. JPG-Bild oder PDF-Datei, das Unterelement „mime type“ enthält eine Angabe hinsichtl. des Datenformats)
dBoolValue	vb	Boolscher Wert
dDescription	dd	Beschreibung (textlich, informell) des Messgeräts/Sensors oder der Systemuhr
device	pd	Informationen zum Messgerät/Sensor
device	tsd	Informationen zur System-Uhr
dFloatValue	vf	Gleitkomma-Wert
dIntegerValue	vi	Ganzzahlen-Wert
distance	dst	Zeitlicher Abstand zwischen den Messwerte (bei isochronen Zeitreihen)
dName	dn	Bezeichnung des Messgeräts/Sensors oder der Systemuhr
dStringValue	vs	Freitext-Wert
exchangeId	xid	Zeitreihen-Bezeichner
exchangeIds	xids	Liste mit einem oder (bei Multiparameter-Zeitreihe) mehreren Zeitreihenbezeichnern
frequency	f	Frequenz mit der (Zwischen-) Messwerte ermittelt werden, angegeben wird der zeitliche Abstand
geoReference	<i>(im Schema nicht als Element modelliert)</i>	Code für das verwendete Koordinaten Bezugssystem, und textlich informeller Ortsbeschreibung

interval	it	Aggregationsintervall, d.h. die Zeitspanne für die der Messwert repräsentativ ist
isochron	iso	Informationen zu Messzeitpunkten mit Startzeitpunkt und zeitl. Abstand (bei isochronen Zeitreihen)
IDescription level	ld l	Ortsbeschreibung (informell, textlich) Schwell- oder Deltawert dessen Über- bzw. Unterschreitung (im Falle einer ereignisgesteuerten Zeitreihe) eine Messwertspeicherung triggert.
IName location	ln dl	Ortsbezeichnung (informeller Name) Ortsbezug für das Messgerät (Messgerät/Sensor oder Systemuhr)
location	tl	Ortsbezug für die Messgröße (für die gesamte Zeitreihe)
location	vl	Ortsbezug für die Messgröße (für einen Messwert der Zeitreihe)
locationDescription	ldn	Informationen zur Ortsangabe mit Name und informeller Beschreibung
measurementInaccuracy (Kindelement von dataValueQuality)	vmi	Unsicherheit des Messwerts
measurementInaccuracy (Kindelement von timeStampQuality)	tsmi	Unsicherheit des Zeitstempels
mimetype	mt	Datenformat einer binären Größe (z.B. für JPEG-Bild oder PDF-Datei)
oDescription	od	Beschreibung (textlich, informell) der Organisation die die Zeitreihe ermittelt hat
offset	ot	Zeitlicher Offset (zur jeweils nächstgrößeren Zeiteinheit) mit der das Messintervall beginnt. Z.B. bedeutet ein Offset von 5min bei einer Intervalllänge von 30min: Das Zeitintervall startet jeweils 5min und 35min nach jeder vollen Stunde.
oName	on	Name der Organisation, welche die Zeitreihe ermittelt hat
organization	org	Informationen zur Organisation/Institution welche die Zeitreihe ermittelt hat
parameter	p	Art der Messgröße (Freitext)
parameter	xp	Art der Messgröße (Eintrag aus Codeliste)
parameterMetaData	pmd	Informationen zum Typ der Zeitreihe
parameterMetaDataList	pmdl	Liste mit Informationen zum Typ der Zeitreihen
point	(im Schema nicht als Element modelliert)	Ortsangabe (als 2D- oder 3D-Koordinaten)
point3D	pt3d	3D-Ortsangabe
pt	pt	2D-Ortsangabe
px	px	Rechtswert-Koordinate
py	py	Hochwert-Koordinate
pz	pz	Höhen-Koordinate
referenceSystem	rs	Koordinatenbezugssystem (Freitext)
referenceSystem	xrs	Koordinatenbezugssystem (Eintrag aus Codeliste)
timedDataElement	tde	Enthält einen oder mehrere Messwerte und optional (falls keine 'isochron-Zeitreihe') einen Zeitstempel
timeSeries	tse	Zeitreihe
timeSeriesList	tseL	Liste mit einer oder mehreren Zeitreihen
timeStamp	sts	Startzeitpunkt der Zeitreihe
timeStamp	ts	Zeitstempel für einen Wert

timeStampMetaData	tsmd	Qualitätsangaben zum Zeitstempel, mit Angabe zur Unsicherheit, Fehlercodes und allgemeinen Angaben zur Systemuhr
timeStampPosition	tsp	Position des Zeitstempels im Zeitintervall (Freitext)
timeStampPosition	xtsp	Position des Zeitstempels im Zeitintervall (Eintrag aus Codeliste)
timeStampQuality	tsq	Qualitätsangaben zum Zeitstempel, mit Angabe zur Unsicherheit und Fehlercode
timeStampQualityRemark	tsqr	Qualitätsbezeichner zum Zeitstempel (Freitext)
timeStampQualityRemark	xtsqr	Qualitätsbezeichner zum Zeitstempel (Eintrag aus Codeliste)
tsValue	tsv	Zeitstempel-Wert
unit	u	Maßeinheit der Messgröße (Freitext)
unit	xu	Maßeinheit der Messgröße (Eintrag aus Codeliste)
UUID	uuid	ID der Zeitreihe
value	<i>(im Schema nicht als Element modelliert)</i>	Enthält einen der folgenden Werte-Objekte: dBoolValue (ja/nein - Aussage), dIntegerValue (Ganzzahl), dFloatValue (Gleitkommazahl), dStringValue (Text), dBinaryValue (binär kodierte Daten), dAnyValue (beliebig selbst definierter Datentyp)
values	vl	Enthält einen oder mehrere Messwerte

Abbildung 8: Gegenüberstellung UML-Bezeichnung zu XML-Schema-Elementname

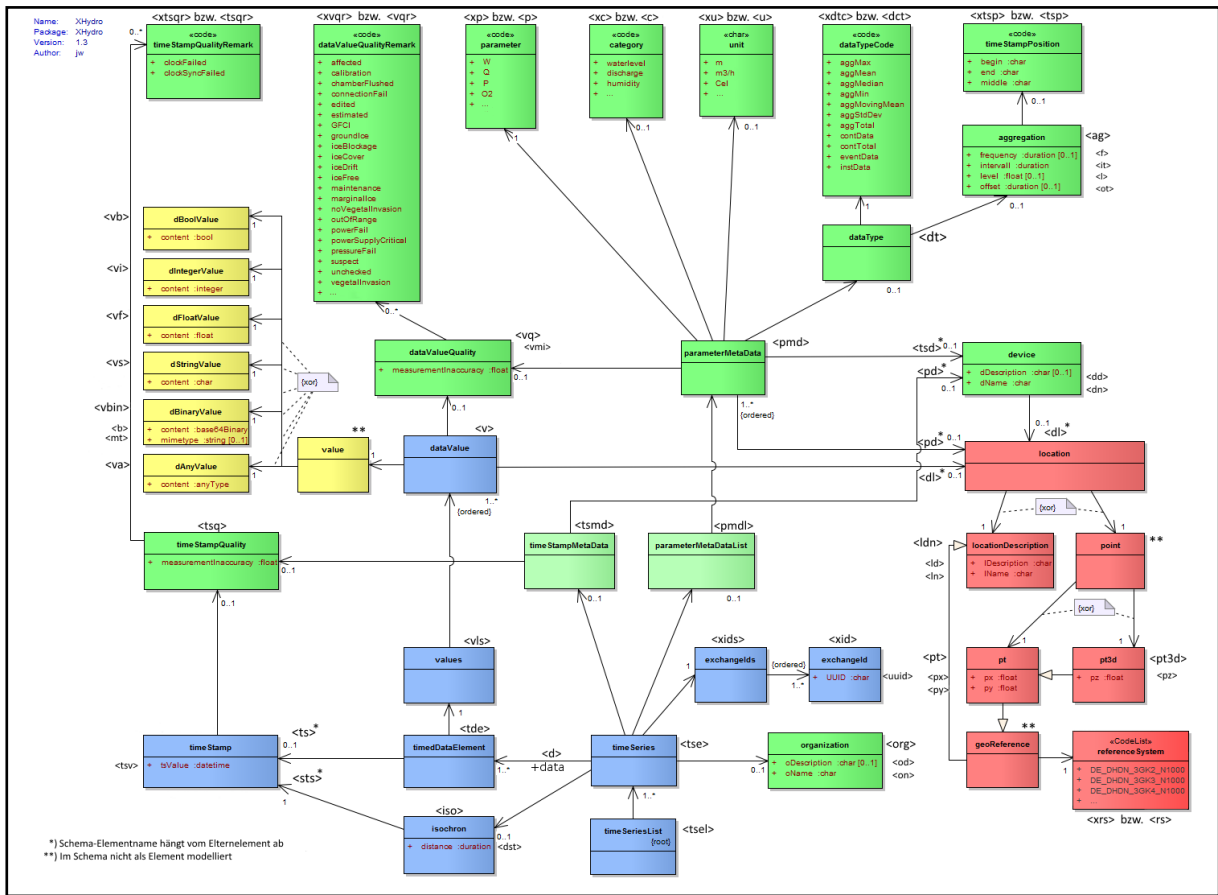


Abbildung 9: UML-Diagramm mit XML-Elementnamen

4 Literatur

- [1] Wikipedia:
XML Schema
https://de.wikipedia.org/wiki/XML_Schema

- [2] Wikipedia:
Unified Modeling Language
https://de.wikipedia.org/wiki/Unified_Modeling_Language

- [3] Bundesanstalt für Gewässerkunde:
Schema Dokumentation
<http://www.xhydro.org/dokumentation.html>

- [4] Wikipedia:
Universally Unique Identifier
https://de.wikipedia.org/wiki/Universally_Unique_Identifier

- [5] Iana.org
Media types
<http://www.iana.org/assignments/media-types/media-types.xhtml>

- [6] Gunther Schadow and Clement J. McDonald:
The Unified Code for Units of Measure
<http://unitsofmeasure.org>

- [7] Wikipedia
Messunsicherheit
<http://de.wikipedia.org/wiki/Messunsicherheit>