

UncertWeb

The *Uncertainty Enabled Model Web*

SEVENTH FRAMEWORK PROGRAMME

THEME FP7-ICT-2009-4

ICT for Environmental Services and Climate Change Adaptation

Deliverable 8.1

Report on integration requirements in UncertWeb

Title of Deliverable	Report on integration requirements in UncertWeb
Deliverable reference number	D8.1
Related WP and Tasks	WP4 (Task 4.1), WP5 (Task 5.1), WP6 (Task 6.1), WP7 (Task 7.1), WP8 (Task 8.1)
Type of Document	Public Report
Authors	Christoph Stasch, Lydia Gerharz, Benjamin Pross, Edzer Pebesma
Date	22/11/2010
Version	1.0

Project coordinator

Dr. Dan Cornford
Aston University, United Kingdom
E-mail: d.cornford@aston.ac.uk

<http://www.uncertweb.org>

Revision History

Version	Date	Changes	Authors
0.0	16/09/2010	Initial template	Edzer Pebesma, Benjamin Pross, Christoph Stasch
0.1	13/10/2010	Section 2 added	Lydia Gerharz
0.2	15/10/2010	Section 3 added	Christoph Stasch
0.3	17/10/2010	Section 4 added	Benjamin Pross
0.4	20/10/2010	Executive summary added, revision of 2,3,4	Edzer Pebesma
0.5	22/10/2010	Finished, up to inclusion of D7.1	Edzer Pebesma
0.6	27/10/2010	Review, added questions, minor modifications to content	Dan Cornford
0.7	05/11/2010	Changes made according to review	Lydia Gerharz, Benjamin Pross, Christoph Stasch
1.0	22/11/2010	Changes made to accommodate D7.1.	Lydia Gerharz, Benjamin Pross, Christoph Stasch, Edzer Pebesma

Related task(s):

Task 4.1 User requirements for developing a tool for the validation of land cover data in African Protected Areas

This task will focus on defining the user requirements necessary to define the means to quickly visually validate high resolution multi-temporal satellite data (Landsat images) using ground truth data within and around African Protected areas. The user requirements will also address the reporting of uncertainties as well as the data discovery and exchange of data between the Web-based land-cover validation tool and the JRC database that is hosting the images. The requirements will be fed into WP1 and WP2 to inform the design of the Model Web encoding and interfaces.

Active partners: JRC, AST

Task 5.1 Defining the application domain requirements for service chaining

This task documents the models to be considered in the work package. This includes a description of each model's inputs and outputs, the computational complexity of the model and how the models link together. The interaction of socio-economic and policy models with the dynamic land-use model is of greatest importance. The final output will be a report that helps inform methodological developments in WP1-3.

Active partners: FERA, AST

Task 6.1 Establishment of data flows, standards and service chaining

This task will interact directly with WP1-3 and WP8 in order to effectively implement the protocols and standards for use in the uncertainty-enabled model Web. Current GMES services will be the main source of data for driving the ensemble forecasts. These source data will include a) ensemble weather forecasts available from ECMWF through GMES; b) ensemble regional scale air quality forecasts available from 7 models through the GEMS

project; c) local near-real time monitoring data, currently available at NILU. The output of the local scale ensemble air quality forecasts will also be addressed using the same standards and methodologies. This task will ensure a standardised implementation of the uncertainty-enabled model. Web linkages, will implement service discovery tools within this use case and will interact directly with WP8 in defining and assessing these tools and links.

Active partners: NILU, AST

Task 7.1 Requirements for creating the uncertainty-enabled model Web

This task will define the requirements for the individual behaviour models; their inputs, outputs and how the uncertainty in these is best represented and propagated in a model Web context.

Active partners: TUE, AST

Task 8.1 Integration requirements evaluation

An overview will be given, based on the experiences of all partners involved in all case studies, as to which information is required for quantitative assessment of uncertainty, how well UncertML adapts to their application domain, and which requirements on information encoding or processing are further required. Input will also be sought from the Project Advisory Board.

Active partners: UOM, AST, JRC, CNR, FERA, WU, NILU, TUE

Legal Notices

The information in this document is subject to change without notice.

The Members of the UncertWeb Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the UncertWeb Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Executive Summary

This report integrates and summarizes the requirements to the UncertWeb project that were described in the requirements document of the individual work packages (D4.1, D5.1, D6.1, D7.1). It summarizes requirements focusing on three aspects: (i) quantitative aspects of uncertainty (see also D1.1), (ii) information encoding, and (iii) geoprocessing.

This report is an inventory of all requirements put forth in the case studies. It does not suggest a prioritization, nor does it evaluate what is realistically realizable within the time frame and budgetary limits of the UncertWeb project.

Section 2 describes how uncertainty is currently assessed, described and represented for model inputs and outputs as well as model error in the application work packages WP4-WP6. The second part of the section lists how uncertainty will be assessed in WP4-WP6, and which information is required for this assessment.

Section 3 describes the requirements for the data encodings that will be used in the scenarios of WP4-WP7. The data inputs, outputs and parameters description of the deliverables D4.1-D7.1 have been analysed and the results are summarized. It details requirements on the spatial encoding (3.1), the temporal encoding (3.2), the observations (3.3) and parameters which are neither spatial nor temporal (3.4).

Section 4 describes the requirements for the web based processing that will take place in the scenarios of WP4-WP7. The requirements concerning the web based processing of the deliverables D4.1-7.1 have been analyzed and described. Technical requirements, including hardware and software needed to execute the models over the web, are discussed.

Section 5 describes the profiles developed for GML and O&M. GML and O&M provide a very wide range of possibilities to encode information, and these profiles define which components of them will be supported by UncertWeb. The profiles themselves form a digital appendix to this report, and are not included in the main text. These files are rather large, and are primarily meant to be machine-readable, rather than human readable.

With this *Report on integration requirements in UncertWeb* the project can proceed to decide upon a ranking of priorities and thus find the most efficient and economic path to realize the uncertainty enabled model web.

Contents

1	Introduction	1
1.1	Definitions	1
1.2	Abbreviations.....	1
2	Information Required for Quantitative Assessment of Uncertainty.....	3
3	Requirements on Information Encoding	1
3.1	Requirements on Spatial Encodings	1
3.1.1	Requirements for raster based encodings.....	1
3.1.2	Requirements for vector based encodings.....	3
3.2	Requirements on Temporal Encodings.....	4
3.3	Requirements on Observation Components	5
3.4	Non-spatial/temporal Data.....	7
4	Requirements on Processing	9
4.1	Requirements on web processing services	9
4.1.1	Adaptation of input/outputs.....	9
4.1.2	Communication patterns (synchronous vs. asynchronous communication)	9
4.2.1	Protocols needed	11
4.2.2	Security constraints	11
4.2.3	Soft- and Hardware requirements	12
4.2.3.1	Wrapping the models	12
5	Annex	14
5.1	GML Profile	14
5.1.1	ISO:19139 Data Quality Extension.....	16
5.2	O&M 2.0 Profile.....	17
5.2.1	Temporal restrictions.....	17
5.2.2	Spatial restrictions	19
5.2.3	ResultQuality restrictions.....	20
5.2.4	Result restrictions.....	20
	References	22

1 Introduction

UncertWeb is a European Project funded under the 7th framework programme. Its main goal is to realize the uncertainty enabled model web. The model web is a concept where the modelling of complex processes which may be spatial, temporal and spatio-temporal takes place by the coupling, or chaining, of individual, interoperable web services that provide either data sources or data processing capability. In order to find out what the quality, or value of an output of a model web is, uncertainties in data and/or models need to be quantified and communicated. The UncertWeb project will develop and apply methods and technology to realize this.

In order to apply the methods and technology developed a number of case studies of existing and typically coupled models have been selected, as case studies in the project. These case studies all have their requirements with respect to the types of variables addressed and the representation of uncertainty within them, the data structures and file formats used, and the technical environment or setup in which they operate. Each of the case studies has produced a deliverable defining these requirements (D4.1, D5.1, D6.1 and D7.1) and this report integrates and summarizes the findings of them.

With this *Report on integration requirements in UncertWeb* the project can proceed to decide upon a ranking of priorities to find the most efficient and economic path to realize the uncertainty enabled model web.

1.1 Definitions

In order to ensure a standard set of terms is used throughout the following definitions are adhered to throughout the document.

Model web / Model chain: a chain of model components connected through web service interfaces.

Model component: a representation of a process or series of processes implemented as computer code, often called a simulator.

Model input: a value, or series of values (if a field), that must be provided to the model component to evaluate the model (likely to include parameters in the model component, initial and boundary conditions for the model component).

Model output: a value, or series of values (if a field), that is produced by the model when it is evaluated.

Observation component: a set of observations of reality that might be used as model inputs, but also might be used to compare against model outputs. The distinction between an observation component and a model input is that an observation component is used to describe observations (which are collected and stored – i.e there is some data there) which might be used in the processing chain somewhere, often for output validation. Model inputs are descriptions of what the models want to get as their inputs, which might not always be directly comparable to the observations we have available.

1.2 Abbreviations

GDAL	Geographic Data Abstraction Layer
GML	Geography Markup Language
OGC	Open Geospatial Consortium

O&M
PDF

Observations & Measurements
Probability Distribution Function

2 Information Required for Quantitative Assessment of Uncertainty

This section describes how uncertainty is currently assessed, described and represented for model inputs and outputs as well as model error in the application work packages WP4-7. The second part of this section lists how uncertainty will be assessed in WP4-7 of UncertWeb, and which information is required for this assessment. This section complements the more technical UncertML requirements, which also considers issues external to the UncertWeb project, developed in D1.1 [1].

2.1 Current quantitative uncertainty assessment

The current uncertainty assessment in the application scenarios of UncertWeb can be described by the uncertainty sources, types, marginal uncertainty representation, joint uncertainty representation, model error, and cross validation methods. Data formats are given together with uncertainty types and representations.

2.1.1 Uncertainty sources for input data

Sources of input uncertainty are not represented in UncertWeb but may help to identify missing information for planned uncertainty quantification (see Section 2.2). Identified uncertainty sources are listed in Table 1:

Source	Description	Example
Interpolation/Aggregation error from point data	Error due to aggregation from point-level data (observations)	CurWeather (WP5)
Misclassification of categorical data	E.g. land use only gives the major land use class per pixel and ignores other land use types	Geophysical Data, TAPM: SURFACE PARAMETERS (WP6)
User input	Classification rules, misinterpretation, non-response	e-habitat:SelectedArea (WP4), LCCS:rules (WP5), Albatross-Activity diary data (WP7)
Sampling error	Error introduced by sampling from a larger population, dependent on sample size	LandSFACTS:Inicrop/Historical Crop Rotation Data: IACS (WP5)
Measurement error	Considered less important, except malfunctioning	Meteorological/Air quality observations (WP6)
Spatial representativeness	For air quality observations the representativeness of the measured value depends on weather conditions and location	Air quality observations (WP6)
Resolution of input data	Uncertainties may be dependent on the scale or resolution of the data (e.g.	e-habitat:slope (WP4)

	slope derived from DEM).	
--	--------------------------	--

Table 1: Uncertainty sources identified in the application scenarios of UncertWeb

2.1.2 Uncertainty types

Uncertainty types describe to which data property the uncertainty refers. The three main uncertainty types are usually referred to as positional (location), attribute and temporal uncertainty. In UncertWeb temporal uncertainty has not been identified so only attribute and locational uncertainty occur as shown in Table 2:

Type	Description	Data type	Data format	Example
Location	Boundaries of protected areas and Extent of Occurrence maps are uncertain	Continuous position Categorical attribute (yes/no)	Vector/Raster data (converted from vector data)	only e-habitat:SelectedArea and e-habitat:EOO/AOO maps for mammals (WP4)
Attribute		Continuous/ Categorical/ Discrete	Vector/Raster/non-spatial data	e-Habitat:PoHS (WP4), LandSFACTS:Inicrop (WP5), LCCS:LCdatasets (WP5)

Table 2: Uncertainty types identified in the application scenarios of UncertWeb

The uncertainty in e-habitat:EOO/AOO maps for mammals is described as locational. However, if this uncertainty about the locations /extents of mammals is represented on a raster framework this would be rather treated as attribute uncertainty.

- RQ1. Attribute uncertainty needs to be supported in UncertWeb for continuous, categorical and discrete data in vector, raster and non-spatial formats.
- RQ2. Locational uncertainty needs to be supported in UncertWeb for continuous vector data.

2.1.3 Uncertainty representations

Uncertainty can be represented by different statistical measures, realisations/samples or fully described by probability distribution functions. In UncertWeb different representations of uncertainty are used for input and output data in the application scenarios as shown in Table 3.

Representation type	Description	Data type	Data format	Example
Ensembles	Limited number of model outputs produced by perturbation of initial conditions or different models	Continuous, categorical	Vector/Raster data	Perturbation of input: TAPM: ECMWF ENSEMBLE (WP6) Different models: EPISODE: REGIONAL BACKGROUND (WP6)
Realisations	A large number of samples produced by statistical simulation or drawn from a PDF	Continuous, categorical	Vector/Raster data	e-Habitat:climate (WP4), LandSFACTS:CropArea (WP5)
Marginal probability distribution function	PDF describing uncertainty for marginals (e.g. for each location)	Continuous	Vector/Raster data	e-habitat:elevation, e-habitat:pTreeCover (WP4)
Joint probability distribution function	PDF describing joint uncertainty (i.e. under consideration of correlation over space/time of one variable)	Continuous, categorical	Raster data	Only partially,error correlation functions: e-habitat:elevation (WP4) Between ensembles: TAPM: ECMWF ENSEMBLE (WP6)
Joint probability distribution function (across multiple inputs)	Joint PDF, describing correlation across different variables	Continuous	Raster data	only possible for elevation data and of minor importance (WP4)
95 % confidence interval	Lower and upper boundaries for 95 % confidence interval	Continuous	Non-spatial data	Expected Yield Data – Defra (WP5)
Index of accordance [%]	% of correctly predicted plots	Categorical	Raster data	e-habitat EOO/AOO maps for mammals (WP4)
% measurement error	Error assessed by quality control or given as single estimate per measurement device type	Continuous	Vector data (points)	Meteorological observations (WP6)
% classification error	Classification for categorical data	Categorical	Raster/Vector data	e-Habitat:landcover (WP5)
Stratified samples	Low, average, high estimates	Continuous	Non-spatial data	Expected Yield Data – Farm Management Pocketbooks (WP5)
Standard error	Standard error of an estimated parameter, e.g. regression parameter	Continuous	Raster data	e-habitat:pTreeCover (WP4)
Standard deviation	Standard deviation from e.g. validation	Continuous	Raster data	e-habitat:elevation (WP4)
R ² from validation	Estimated by comparison with validation data, represents percentage of explained variability	Continuous	Raster data	e-habitat:pTreeCover (WP4)

Table 3: Uncertainty representations used in the application scenarios of UncertWeb

Marginal uncertainty is often not quantified or only partially assessed by statistical measures in the case studies. A full description based on a probability distribution function seems to be possible for some variables, but it is unclear if this has already happened. Joint uncertainty is most often unknown and not planned to be assessed except for e-Habitat:pTreeCover, e-Habitat:elevation, e-Habitat:landcover input for WP4. Joint uncertainty across multiple variables has not been assessed so far and is most often considered less important. However, joint uncertainty will be essential in model outputs and maybe necessary to consider for visualisation and spatio-temporal aggregation of the results.

Realisations or ensembles are frequently used in at least two case studies (WP5, WP6). Besides this, a larger number of representation types are only used for one data set in a single case study specifically. It is not yet clear how far these uncertainty quantifications are used in the uncertainty propagation for the model output and thus if, and to what extent they need to be supported by UncertWeb.

- RQ3. Realisations need to be supported in UncertWeb for continuous as well as categorical raster and vector data.
- RQ4. Marginal and joint probability distribution functions need to be supported in UncertWeb for continuous raster and vector data.
- RQ5. A variety of different statistical measures are used for uncertainty representation in UncertWeb and should be considered for support – see D1.1.
- RQ6. Joint uncertainty in model outputs needs to be supported in UncertWeb.

2.1.4 Model error

Besides the input error, the model error is potentially introduced by each processing step and is an important measure for uncertainty quantification in the output. Currently, model errors are not quantified in UncertWeb and are not planned to be improved beyond the current state.

2.1.5 Cross validation

Cross-validation of model outputs can be used to quantify uncertainties in the outputs. The current availability of observations and applied cross validation methods are listed in

WP7	Albatross-Activity diary data	Raw observations of individual activities	Continuous	Non-spatial data	Albatross-Time dependent activity schedule (WP7)
WP7	Traffic flows observations	observations of traffic flows for a number of road segments	Continuous	Vector data (road segments)	MADAM- Traffic flow condition

Table 4:

Work package	Observations available	Description	Data type	Data format	Cross validation of outputs
WP4	Probably not	N/A	N/A	N/A	Currently no validation method, might be possible for species habitats
WP5	Historical Crop Rotation	Historical data for field	Continuous, Categorical	Non-spatial	LandSFACTS:CropArea,

	Data	samples from IACS and June Survey Data		data	LandSFACTS:SimRot (WP5)
WP6	Meteorological and Air Quality observations	Limited number of fixed measurement stations in Oslo	Continuous	Vector (point) data	TAPM: ENSEMBLE OUTPUT, EPISODE:ENSEMBLE OUTPUT (WP6)
WP7	Albatross-Activity diary data	Raw observations of individual activities	Continuous	Non-spatial data	Albatross-Time dependent activity schedule (WP7)
WP7	Traffic flows observations	observations of traffic flows for a number of road segments	Continuous	Vector data (road segments)	MADAM- Traffic flow condition

Table 4: Cross validation methods used in the application scenarios of UncertWeb

2.2 Required information for planned uncertainty quantification

One main objective of UncertWeb is to enable models to handle uncertain inputs as well as quantify uncertainties in the outputs. The planned activities to assess uncertainty beyond the current state in the case studies it shown in tab 5:

Planned uncertainty quantification method	Description	Uncertainty representation	Information required	Example
Expert elicitation	Eliciting uncertainty estimates from experts	varying	Expert knowledge, review of methods	e-habitat:SelectedArea (WP4), Expected Yield Data – FERA (WP5), Air quality observations (WP6), Albatross-Activity diary data (WP7)
Modelling with Dirichlet distributions	unknown	Transition probabilities matrices	unknown	LandSFACTS:Tmats (WP5)
Quality control	Further quality control of measurements	% measurement error or failure	Quality information	Meteorological observations (WP6)
Interpolation error	Error from interpolation of point	Standard deviation or normal PDF	Original data	e-Habitat:climate (WP4), CurWeather (WP5)

	measurements			
Transfer results from other assessments	Use uncertainty assessment results from similar studies	varying	Review of assessments, expert knowledge for transfer	TAPM: ENSEMBLE OUTPUT, EPISODE:ENSEMBLE OUTPUT (WP6)
Approximate PDF from R ²	Approximate full PDF from R ² of regression parameters	PDF	unknown	e-habitat:pTreeCover (WP4)

Table 5: Information required for further uncertainty assessment in the application scenarios of UncertWeb

These detailed activities can be summarised to three general approaches that are relevant for UncertWeb:

- RQ7. Expert elicitation needs to be supported in UncertWeb for different domains and data sources.
- RQ8. Capturing the results of cross-validation using data internal to the project needs to be supported by UncertWeb.
- RQ9. Refining estimates of uncertainty through quality control procedures and the use of external assessments needs to be supported by UncertWeb.

2.3 Requirements for uncertainty visualisation

The outputs of the applications scenarios should be visualised if possible with the associated uncertainties. The requirements for a visualisation tool to support all current and planned uncertainty quantifications are listed for input variables in Table 6 and for output variables in Table 7.

Uncertainty type	Data type	Data format	Uncertainty representation	Description	Example
Location	Continuous data	Vector data	pdf	Uncertainty of the polygon boundaries	e-habitat:SelectedArea (WP4)
Attribute	Continuous data	Raster data	pdf	Has to be quantified	e-habitat:pTreeCover (WP4)
Attribute	Continuous data	Raster data	Standard deviation	From validation	e-habitat:elevation (WP4)
Attribute	Continuous data	Raster data	Realisations	Realisations from point interpolation	e-Habitat:climate (WP4)
Attribute	Continuous data	Raster data	Ensembles	Assessed by perturbation of input parameters	TAPM: ECMWF ENSEMBLE, EPISODE: REGIONAL BACKGROUND (WP6)
Attribute	Categorical data	Raster data	Classification error [%]	Percentage of wrongly classified pixels	e-Habitat:landcover (WP4)
Attribute	Categorical data	Raster data	Ensembles	Assessed by perturbation of input parameters	TAPM: ECMWF ENSEMBLE (WP6)
Attribute	Categorical data	Raster data	unknown	Uncertainty of occurrence	e-habitat:EOO/AOO maps for mammals (WP4)
Attribute	Continuous data	Vector data	unknown	Uncertainty will be assessed by expert elicitation	Albatross- Activity diary data (WP7)

Table 6: Requirements for uncertainty visualisation of inputs in the application scenarios of UncertWeb

Uncertainty type	Data type	Data format	Uncertainty representation	Description	Example
Attribute	Continuous data	Raster data	unknown	Has to be quantified	e-Habitat:PoHS (WP4)
Attribute	Continuous data	Raster data	Ensembles	Assessed by limited number of ensembles	TAPM: ENSEMBLE OUTPUT, EPISODE:ENSEMBLE OUTPUT (WP6)
Attribute	Continuous data	Vector data	Realisations	Assessed by stochastic model structure	LandSFACTS:CropArea (WP5)
Attribute	Categorical data	Vector data	Realisations	Assessed by stochastic model structure	LandSFACTS:SimRot (WP5)
Attribute	Continuous	Non-spatial	unknown	Has to be quantified	e-Habitat:HRI (WP4), CAM:CTM (WP5)

Table 7: Requirements for uncertainty visualisation of outputs in the application scenarios of UncertWeb

With regard to the types of uncertainties that need to be supported, attribute uncertainty needs to be supported in the tool, as well as positional uncertainty, provided that the inputs for WP4 need to be visualised. The more specific visualisation needs are in the form of maps (4, 5), bar charts (5), graphics (5), gridded or time series data for individual points (6), for different outputs of the Work Packages 4, 5 6 respectively. Furthermore a possibility to visualise statistical summaries of the model outputs and temporal animation (e.g. WP7) will be required.

- RQ10. Raster and vector data visualisation as well as visualisation of non-spatial data, summary statistics and time series needs to be supported in UncertWeb.
- RQ11. Visualisation of uncertain continuous and categorical variables needs to be supported in UncertWeb.
- RQ12. Visualisation of realisations/ensembles, PDFs, standard deviation and classification error need to be supported or translated into a common representation in UncertWeb.

3 Requirements on Information Encoding

This section describes the requirements for the data encodings that will be used in the scenarios of WP4-WP7. Therefore, the data inputs, outputs and parameters description of the deliverables D4.1-D7.1 [2,3] have been analysed and the results are described below. We distinguish between requirements on the spatial encoding (3.1), the temporal encoding (3.2), the observations (3.3) and the data which is neither spatial nor temporal and usually are parameters based on basic data types (3.4).

3.1 Requirements on Spatial Encodings

In the scenarios of UncertWeb raster as well as vector based encodings are used as model inputs or outputs. In a raster representation, space is divided into an array of rectangular cells. One or more variable values are attached to these cells. Raster based encodings are usually used for representing field based phenomena. Instead, discrete objects in space are usually encoded as vector data. In a vector representation, lines or polygons are captured as a collection of points connected by (usually straight) lines.

3.1.1 Requirements for raster based encodings

The raster formats which are used in the scenario are shown in Table 8. There are a variety of raster formats used in the current models. Most of the formats are supported through standard tools like Geospatial Data Abstraction Library (GDAL, <http://www.gdal.org/>).

Name of the Format	Description	Examples
GeoTIFF	TIFF variant with embedded georeferencing information	e-habitat:PoHS – coverage containing the probability of habitat similarity for each pixel (WP4)
HGT (NASA Height Data)	NASA specific format for raw height values encoded as 16-bit integers; the georeferencing information is contained in the name of the files, e.g. “N34W119.hgt”; HGT stands for “height” and is not a format definition	e-habitat:elevation – height data used as input to the e-habitat model (WP4)
ADF (Arc/Info Binary Grid Format)	Arc/Info Binary Grid Format; converters are available from ADF to TIFF; JPEG, PNG, BMP	e-habitat:waterBodies - percentage of area covered by water bodies (WP4)
BIL (Band Interleaved by Line)	Multi-band binary image format usually used for satellite images	e-habitat:landcover – 27 categories of landcover (WP4)
netCDF (network Common Data Form)	Machine-independent binary data formats that support the creation, access, and sharing of array-oriented scientific	CurWeather - Current temperature and rainfall data provided by UK MetOffice (WP5)

	data.	
GRIB1/GRIB2 (GRIdded Binary)	Bit-oriented data exchange format defined by the World Meteorological Organization (WMO)	TAPM:ECMWF ENSEMBLE - Large scale synoptic meteorology include wind, temperature, water vapour and pressure as 3D fields. (WP6)

Table 8: Raster formats used in the application scenarios of UncertWeb

The types of the variables in the raster data sets are shown in Table 9. They are either continuous or categorical.

Type	Description	Example
Continuous	Continuous values usually encoded as numerical float values. This is the data type of most variables encoded as raster data.	CurWeather - Current temperature and rainfall data provided by UK MetOffice (WP6)
Categorical	Categories which are usually encoded as integer values. The mapping from categories to integers is usually not contained in the coverages. In some cases, it is provided in external files, e.g. Excel spreadsheet for landcover data.	e-habitat:landcover – 27 categories of landcover (WP4)

Table 9: Types of variables in coverages

The raster layers are provided in different resolutions. These are listed in Table 10. There are different resolutions ranging from 30m up to 0.25x0.25 degrees.

Resolution	Example
1km ² Grid/30arc-seconds	e-habitat:pTreeCover (WP4)
30m Grid	e-habitat:elevation (WP4)
0.25x0.25 degrees with 9 vertical levels	TAPM:ECMWF ENSEMBLE (WP6)
25km resolution	EPISODE:REGIONAL BACKGROUND (WP6)

Table 10: Different resolutions of raster data in UncertWeb

As the spatial resolution and the extent of the coverages varies, the file sizes vary as well and range from very small datasets, e.g. 7KB for CurWeather, up to very large datasets, ~900MB for e-habitat:pTreeCover.

Resulting from the tables shown above, the following requirements can be listed for raster based encoded:

- RQ13. Several different data formats need to be integrated and thus supported in UncertWeb. These are listed in Table 8. If NetCDF is going to be used as the only uncertainty-enabled binary encoded coverage format, data converters are needed to integrate the different formats into the infrastructure. Besides that there is a need for linking the uncertainty information to the common raster formats.

RQ14. Continuous as well as categorical variables need to be supported in UncertWeb

RQ15. UncertWeb has to provide mechanism for integrating large (uncertain) coverages with varying resolutions.

3.1.2 Requirements for vector based encodings

The vector formats which are used in the scenario are shown in Table 11. There are two de-facto standards used defined by ESRI Inc. These are the shape format and the geodatabase format. Furthermore, there are proprietary formats like Excel spreadsheets containing spatial information as latitude and longitude columns.

Name of the Format	Description	Example
SHP	De-facto standard defined by ESRI for encoding vector data.	e-habitat:SelectedArea (WP4)
GDB	ESRI Geodatabase format. Database used in ESRI ArcInfo to store vector-based information.	e-habitat:EOO-maps (WP4)
Proprietary Encodings	Data is sometimes contained as lat long in Excel spreadsheets or external sources.	Historical Crop Rotation Data: IACS (WP5)

Table 11: Formats of vector data used in the UncertWeb scenarios

The type of variables encoded as Vector data are shown in Table 12. Besides continuous and categorical variables which are also encoded in raster formats, also Boolean variables are encoded in vector formats.

Type	Description	Example
Continuous	Continuous values usually encoded as numerical float values.	Geophysical Data - pH Value from the LandIS - Land Information System – held and maintained by The National Soil Resources Institute (NSRI) at Cranfield University. The database covers the areas of England and Wales. (WP5)
Categorical	Categories which are usually encoded as integer values. The mapping from categories to integers is usually not contained in the coverages. In some cases, it is provided in external files, e.g. Excel spreadsheet for landcover data.	LCCS:LCfieldalloc - This output of the Land Classification System of FERA contains a land capability classification for each field parcel within the region of interest. (WP5)
Boolean	Either true or false.	e-habitat:EOO-maps – extent of occurrence maps; animals exist (1) or not (0) in a certain region.(WP4)

Table 12: Types of variables in vector data

The geometry types which occur in the UncertWeb scenarios are listed in Table 13. The three basic geometries (Point, LineString, Polygon) are used as well as collections of them (Multi-Polygon).

Type	Description	Example
MultiPolygon	Collection of Polygons which might have also interior boundaries. Thus, the polygon might have one or more exterior boundaries as well as interior boundaries. The boundaries are a set of points, connected through straight lines.	e-habitat:SelectedArea - area for which the indicators are assumed to be optimal in the biodiversity model (WP4)
Polygon	Polygon defined through one or more exterior boundaries as well as interior boundaries. The boundaries are a set of points, connected through straight lines.	LandsFACTS:CropArea (WP5)
LineString	Non-closed selection of points connected through straight lines.	MIMOSA - Traffic flow (WP7)
Point	Point location.	Air Quality Observations (WP6)

Table 13: Geometry Types used in UncertWeb

Resulting from the tables shown above, the following requirements are defined for vector based encodings in UncertWeb. It is planned to use a profile of the Geography Markup Language (GML) [6] as common format for vector encodings. Thus, these requirements can be considered to be requirements for the profile. A preliminary version of the profile is contained in Annex 5.1.

RQ16. The integration of vector data in SHP files, GDB files and proprietary formats needs to be supported by UncertWeb. Standard tools support the conversion of SHP and GDB files into GML files. It is not possible to write common encoders/parsers for proprietary formats. Thus, a tutorial is needed how to encode the data with GML.

RQ17. Continuous, categorical and Boolean variables need to be supported in vector encodings.

RQ18. The geometries that need to be encoded in vector formats are Point, LineString, Polygon, and MultiPolygon. The Polygon needs to support interior and exterior boundaries.

3.2 Requirements on Temporal Encodings

Temporal information needs to be encoded in the observation components as well as in the model outputs. The types needed are shown in Table 14.

Temporal Type	Resolution	Examples
---------------	------------	----------

Time Stamp	Annual information, daily information, hourly information, every minute information	Historical Crop Rotation Data: IACS (WP5), Albatross – Activity Diary data (WP7)
Time Period	Year, day, hour, minute	e-habitat:NDVI (WP4), CurWeather (WP5)
Time Series	Time Series of years, Time Series of hourly values and every minute values	LandsFACTS:SimRot (WP5)

Table 14: Types of Temporal Information needed in UncertWeb

The temporal information can be distinguished in two different types: A Time Stamp represents a single location in time, whereas a Time Period represents a duration in time defined by a begin point and an end point which are usually Time Stamps. The Time Stamp information has different resolutions like year, day or hour. Similarly, the Time Periods have durations of year, day or hour. Thus, the requirements for the encoding of temporal information are as follows:

- RQ19. The temporal encodings have to support TimeStamps, TimePeriods and TimeSeries.
- RQ20. The temporal resolutions which need to be supported are year, day, hour, and minute.

3.3 Requirements on Observation Components

In UncertWeb, the observation model defined in the Observation & Measurements (O&M) specification [4] of the Open Geospatial Consortium (OGC) shall be used for the encoding of real-world observations as well as for uncertain in- or outputs with spatial and/or temporal support. The basic model is shown in Figure 1.

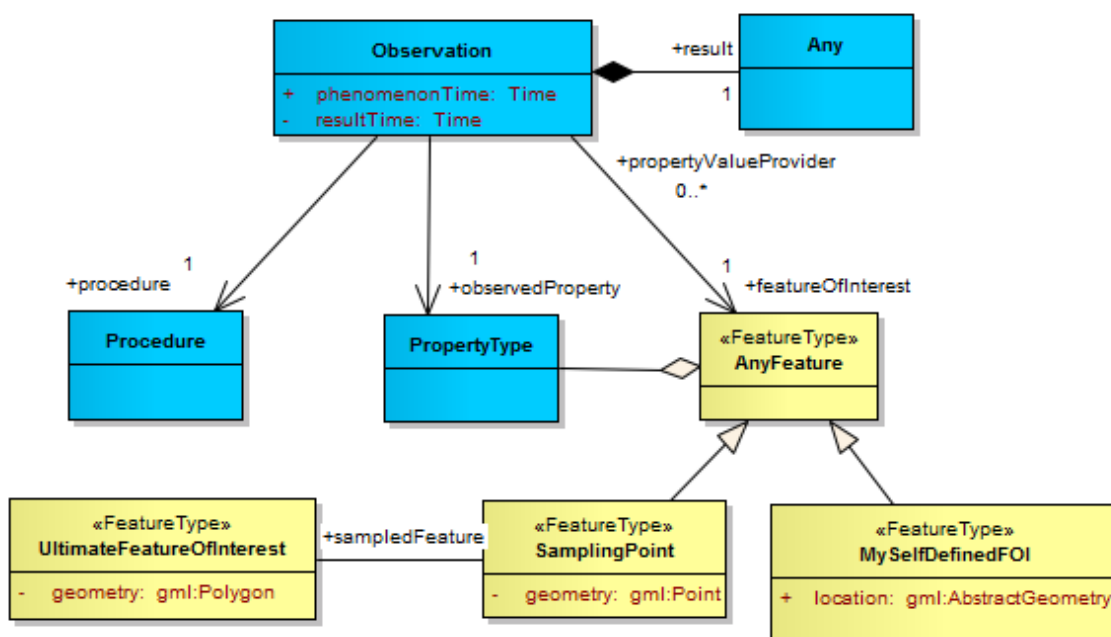


Figure 1: Basic Observation Model

An Observation consists of six basic properties. The procedure element contains information about the sensor or process which has made the observation. The observed property element references information about the property (real world phenomenon), e.g. air temperature,

which has been observed. This information is usually provided in a dictionary. The result value can be anything representing the observation's result like 10°C, a video stream or a statement like “warm”.

The elements containing the spatial information are shown in yellow colour in Figure 1. An observation carries the spatial information in the feature of interest which represents the real world feature for which the observation is providing a value. This can be a human, a building, a lake or something else. In the case of environmental modelling, the observed phenomena are usually spatial fields. Thus, these phenomena are sampled and the feature of interest then is a SamplingPoint containing the concrete location of the observation and a link to the ultimate feature of interest which might be a region where the phenomenon is sampled.

An observation contains two temporal attributes. The phenomenonTime is the time when the observation's result value applies to the feature of interest property. The resultTime represents the time when the observation value is derived, which may be later e.g. in case of chemical analysis in a laboratory.

In the following, we summarise the types of observation components needed in UncertWeb. We derive this information from the description of observation components in D4.1-6.1. The different types of Geometries are shown in Table 15. Polygon and Point geometries need to be supported in the observation components of UncertWeb. The types are further explained in section 3.1.2 above.

Geometry Type	Example
Polygon	EOO-maps: Extent of occurrence maps used in Biodiversity scenario. (WP4)
LineString	MIMOSA – Traffic flow (WP7)
Point	Air Quality Observations: Data provided by 7 air quality stations in Oslo. (WP6)

Table 15: Geometry Types used in Observations

The temporal attributes of the observation components are summarised in Table 16. The observations contain either TimeStamps with hourly or yearly resolution or TimeSeries of hours or years.

Temporal Type	Resolution	Examples
Time Stamp	Hour	CurWeather (WP5)
TimeStamp	Year	Historical Crop Rotation Data: IACS (WP5)
TimeSeries	TimeSeries of years	LandsFACTS:SimRot (WP5)
TimeSeries	TimeSeries of hourly values	Meteorological Observations (WP6)

Table 16: Temporal Types in the Observation Components

As shown in Figure 1 the result type is not further restricted in the result model. Table 17 shows the types of result values of the observation components in UncertWeb. Continuous, Categorical and Boolean result values are used in the observation components of the UncertWeb model chains.

Type	Example
Continuous	Geophysical Data - pH Value from the LandIS - Land Information System – held and maintained by The National Soil Resources Institute (NSRI) at Cranfield University. The database covers the areas of England and Wales. (WP4)
Categorical	LCCS:LCfieldalloc - This output of the Land Classification System of FERA contains a land capability classification for each field parcel within the region of interest. (WP5)
Boolean	e-habitat:EOO-maps – extent of occurrence maps; animals exist (1) or not (0) in a certain region. (WP4)

Table 17: Result Types of Observation Components

Based on the tables shown above, the following requirements are defined for observation components. In UncertWeb, O&M shall be used to encode real-world observations such as air quality measurements or meteorological observations. Additionally, it shall be used for uncertain model in- and outputs with spatial and/or temporal support. Thus, the requirements of this section form the basis for a profile of O&M developed in UncertWeb. A preliminary version of the profile is attached in Annex 5.2.

- RQ21. Observations shall support Polygons, LineStrings and Points as geometries.
- RQ22. Timestamps with hourly or yearly resolution as well as time series of years and hours have to be supported.
- RQ23. Continuous, categorical and Boolean result types need to be supported in UncertWeb.

3.4 Non-spatial/temporal Data

Some model parameters used as model input contain neither spatial nor temporal information but are only based upon basic data types. These are listed in Table 18.

Data Type	Description	Example
Categorical	Category of a common vocabulary.	LCCS:rules - Types of land classifications (WP5)
Transition Matrix	Set of transition matrices with continuous values from [0,1].	CAM: CTM - Transition Matrices for crop rotation (WP5)

Continuous	Continuous numerical values.	Economic Data – forecasts for yields and crop prices. (WP5)
------------	------------------------------	-------------------------------------------------------------

Table 18: Data types of non-spatial/temporal data

Thus, the requirements for non-spatial/temporal data are:

- RQ24. Categorical and Continuous values need to be supported for non-spatial/temporal data types.
- RQ25. Matrices have to be supported.

4 Requirements on Processing

This section describes the requirements for the web based processing that will take place in the scenarios of WP4-WP7. The requirements concerning the web based processing of deliverables D4.1-7.1 [2,3] have been analyzed and the results are described below. Technical requirements like hardware and software needed for executing the models over the web are discussed.

4.1 Requirements on web processing services

From WPs 4.1 – 7.1 three different complexities concerning the transformation of models into web services can be derived. In the e-Habitat model (WP4) there is one model that should be transformed into a web service. Regarding the air quality forecast model (WP6) there is the question whether to link the two underlying models or to treat them as a single model component i.e. process. In the land use scenario (WP5) there are four model components that are subject to transformation into web services.

4.1.1 Adaptation of input/outputs

In Section 3.1 the existing input and output types are already listed. After deciding what the corresponding data types in UncertWeb will look like, a WPS profile that holds information about the supported input/output formats can be defined. According to the WPS specification (OGC 05-007r7), a WPS (Application) Profile should be defined as a standalone document containing:

- An Universal Resource Name (URN) that uniquely identifies the process,
- A reference response to a DescribeProcess request for that process (reference process schema),
- A human-readable description of the process and its implementation (optional, but recommended),
- A Web Service Description Language (WSDL) document for that process (optional).

RQ26. Define a WPS profile for use in UncertWeb

4.1.2 Communication patterns (synchronous vs. asynchronous communication)

The following table recapitulates the information about the estimated computing time of the model components described in the deliverables D4.1-D7.1:

Model Process	Estimated computing time
e-Habitat (PoHS) (WP4)	Slow when modelling 1*1 km ² grid, which is aimed for.
The Land Capability Classification System (LCCS) (WP5)	Dependent on the number of land capability classifications used and the size of the historical crop rotation data set. For the East Anglian Chalk JCA region, computation time is minimal.

The Crop Allocation Mechanism (CAM) (WP5)	Dependent on the number of land capability classifications used and the size of the historical crop rotation data set. For the East Anglian Chalk JCA region, computation time is minimal.
LANDcape Scale Functional Allocation of Crops Temporally and Spatially (LandSFACTS) (WP5)	Dependent on the number of simulations being run and on the area being simulated. The program took about 5 hours to run for 1,000 simulations for an application similar to the ones that will be run in UncertWeb.
Yield Model (YIELD) (WP5)	Computation time is minimal, as this model is not complex.
TAPM-EPISODE (WP6)	TAPM: Typical one day simulation will take around 5 -20 minutes, dependent on resolution. Tradeoffs usually required in terms of resolution. If we consider 100 ensemble members then the total CPU time on a normal PC will be around 33 hours. 8 processors will allow for run times of around 4 hours. EPISODE: For Oslo, one days simulation of air quality including all sources will take around 10 minutes
Albatross (WP7)	Dependent on number of input and Scenario Agents (of whom the simulated population consists).

Table 19 Estimated computational time of the model components

Table 19 shows that some model components take a substantial amount of time to compute. Executing these models via a web service synchronously could lead to problems like timeouts. To avoid that, a mechanism for asynchronous execution is required. On the one hand pull based solutions exist, where the client requests the status of the execution periodically from the service. On the other hand there is the push based solution, where the server notifies the client when the execution of a process has finished. We will have to decide where we will execute a model component synchronously or asynchronously and which mechanism will be most useful for the different processes.

RQ27. Support asynchronous execution of processes/model components, and choose between push- and pull-based mechanism

4.2 Technical requirements

4.2.1 Protocols needed

As mentioned above the services must be accessible through the web. Some partners may have protection mechanisms for their network in place like firewalls or proxy servers that can block communication between the services. The SOAP protocol provides a way to communicate between applications running on different operating systems, with different technologies and programming languages.

4.2.2 Security constraints

The following table lists the datasets used as model input and their availability and current accessibility extracted from deliverables D4.1-D7.1:

Dataset Name	Availability	Accessibility (currently)
e-Habitat:EOO-maps (WP4)	Free ¹	Over the web
e-habitat:SelectedArea (WP4)	Restricted ²	Over the web
e-habitat:pTreeCover (WP4)	Free	Over the web
e-habitat:elevation (WP4)	Free	Over the web
e-habitat:slope (WP4)	Free	Over the web
e-habitat:climate (WP4)	Free	Over the web
e-habitat:NDVI (WP4)	Free	Over the web
e-habitat:waterBodies (WP4)	Free	Over the web
e-habitat:landCover (WP4)	Free	Over the web
UKCP09 projections (WP5)	Restricted ²	Over the web
CurWeather (WP5)	Restricted ²	Over the web
Geophysical Data (WP5)	Restricted	Stored locally
Historical Crop Rotation Data: IACS (WP5)	Restricted	Stored locally
Historical Crop Rotation Data: June Survey (WP5)	Restricted	Stored locally
Economic Data (WP5)	Free	Stored locally
Geometry of Fields Data (WP5)	Restricted	Stored locally
Expected Yield Data - Farm Management Pocketbooks (WP5)	Free	Stored locally
Expected Yield Data – Defra (WP5)	Restricted	Stored locally
Meteorological observations (WP6)	Free	Stored locally
Air quality observations (WP6)	Free	Over the web

Table 20 Availability and accessibility of input datasets

¹ free for scientific use

² users must register prior to access

A part of the datasets will have to be made available via OGC web services. We will have to decide which partner hosts which datasets. As some data is restricted it has to be secured, that no unauthorized person can access or use the data. Also the respective partner must ensure that the services can be accessed over the web and are not blocked e.g. by a firewall.

- RQ28. Make appropriate datasets available through OWS
 RQ29. Restrict access to datasets where necessary

4.2.3 Soft- and Hardware requirements

The following table lists some Soft-and Hardware requirements extracted from deliverables D4.1 – D7.1:

Model Name	Software	Currently used Hardware/OS	Hardware upgrade planned/useful
e-Habitat (PoHS) (WP4)	R	NA	Upgrade may be useful.
The Land Capability Classification System (LCCS) (WP5)	R	Desktop PC/Windows	Upgrade may be useful when investigating larger regions.
The Crop Allocation Mechanism (CAM) (WP5)	R	Desktop PC/Windows	Upgrade may be useful when investigating larger regions.
LANDscape Scale Functional Allocation of Crops Temporally and Spatially (LandSFACTS) (WP5)	LandSFACTS (C++)	Desktop PC/Windows	Upgrade may be useful when running more simulations or investigating wider area.
Yield Model (YIELD) (WP5)	R	Desktop PC/Windows	No need as computation time is minimal.
TAPM-EPISODE (WP6)	TAPM-EPISODE	Desktop PC/Windows	Planning to use multiple processor system.
Albatross (WP7)	Albatross	Desktop PC/Windows	Upgrade may be useful.

Table 21 Information about hard- and software behind the model components

The models used in the scenarios can be executed successfully on a desktop pc with windows as operating system. Executing some complex models on standard machines however can result in long computing times (see Table 19). NILU already plans to use a multiple processor system. This might be useful for the other partners as well in order to reduce computing time.

4.2.3.1 Wrapping the models

In order to access the model via a web service, we need to wrap the model. This is only possible, if the software-framework behind the model can be executed by the web service. LandSFACTS, TAPM and EPISODE are still behind a window-based front-end. It will be required to access the underlying code in order to execute the models via a web service.

Software	Command line or window based	Additional Comments
LandSFACTS	Window based	Source Code (C++) is available.
R	Command line based	WPS wrapping for R is currently under construction
TAPM-EPIISODE	TAPM: Window based EPIISODE: Window based	TAPM: TAPM_GUI.EXE uses batch files for the execution of the model. EPIISODE: Programmed in FORTRAN.
Albatross	NA.	Written in C++.

Table 22 Additional information about software behind model components

5 Annex

This annex contains the preliminary profiles of GML [4] and O&M [6] which will be used in UncertWeb. Following the new OGC guideline of writing modular specifications [5] each specification defines a set of requirements grouped into certain requirements classes. For each requirement class there has to be a conformance test class defining several conformance tests to prove whether an implementation satisfies the requirements.

A *profile* is defined as a “specification or standard consisting of a set of references to one or more base standards and/or other profiles, and the identification of any chosen conformance test classes, conforming subsets, options and parameters of those base standards, or profiles necessary to accomplish a particular function”. In other words, a profile restricts the different options of a base standard to be suitable for a certain application purpose. An implementation which is conformant to a certain profile is always also conformant to the base standard referenced in a profile.

The GML Profile restricts the types to be used from GML to a few very basic geometry types as described in the requirements above. The GML Simple Features Profile [7] is similar to the profile defined in UncertWeb, though the profile defined in UncertWeb is more restrictive. The GML profile is then used in the O&M profile to restrict the possibilities for encoding geometries.

In UncertWeb, the Observations&Measurements format [4] will be used as one format to exchange observations and uncertain results. As the basic specification is quite generic in the way how spatial and temporal positions of observations can be encoded, a simple profile of the observations for encoding spatial and temporal positions will help to ease implementing several software components. This might at the same time become a valuable contribution to the Sensor Web Enablement (SWE) community.

5.1 GML Profile

The profile defined below restricts the geometries to Point, LineString, Polygon, (Rectified-) Grid and MultiGeometry. Several minor restrictions have been made to different possibilities for expressing the same information. For example, in GML 3.2.1 the `gml:Envelope` can be defined by either two elements named `lowerCorner` and `upperCorner` or by two `gml:pos` elements or by a `gml:coordinates` element. This has been changed in the profile to just `lowerCorner` and `upperCorner`. The schema is defined in `Profiles/GML/UncertWeb_GML.xsd`¹. All changes have been made through either changing the elements which are not used to comments or adding an additional comment describing the changes. Figure 2 shows an overview on the `gml:AbstractFeature` defined in the profile. As stated above, the `gml:location` has to carry a subtype of `gml:AbstractGeometry` which can only be `gml:LineString`, `gml:Polygon`, `gml:Point`, `gml:(Rectified-)Grid`, or `gml:MultiGeometry`. The `gml:RectifiedGrid` is a grid for which there is an affine transformation between the

¹ The current version of the schema files and examples are attached in the profiles.zip file.

grid coordinates and the coordinates of an external coordinate reference system. The Grids defined in GML can be used, if the support of e.g. interpolated values is not the cell size of a raster, but the point within a grid. If the Grid is irregular, a `gml:MultiGeometry` containing `gml:Points` can be used.

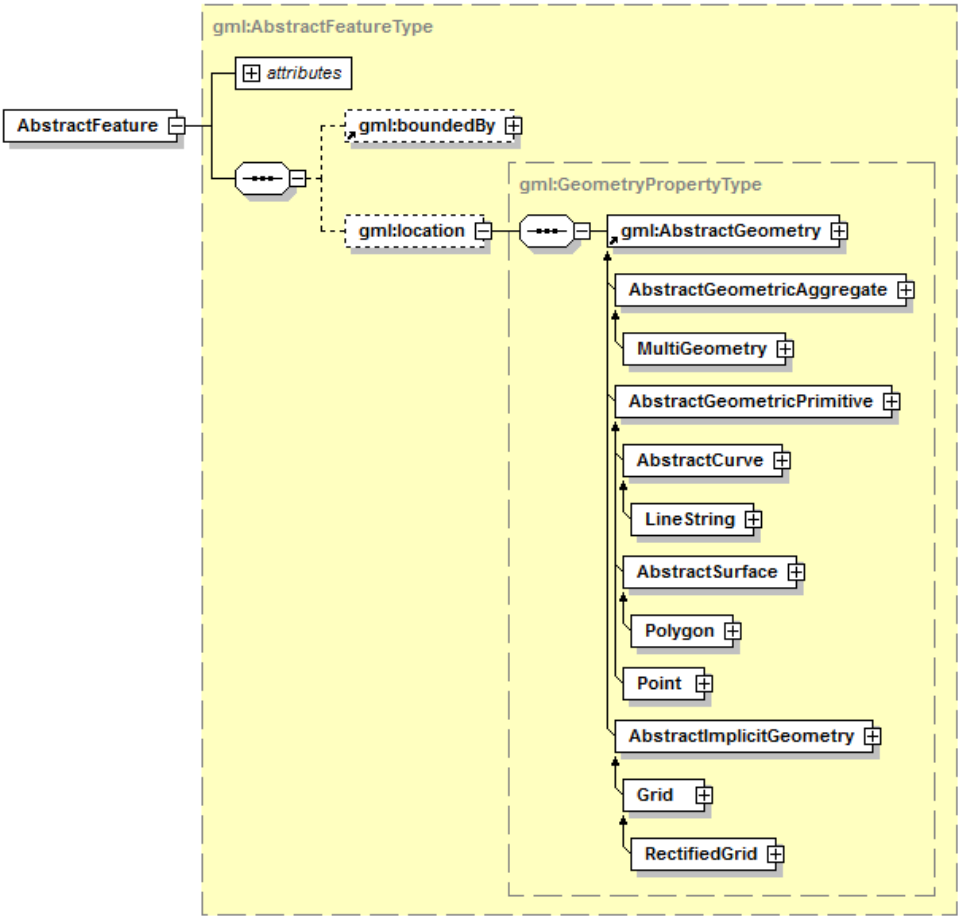


Figure 2: Schema diagram for AbstractFeatureType defined in the UncertWeb GML Profile

We also added an `gml:UncertaintyPropertyType` which can be used for feature properties carrying uncertainty information. The type contains an UncertML uncertainty element plus optional information about the unit of measure encoded as `uom` attribute. Figure 3 shows the schema diagram of the `gml:UncertaintyPropertyType`.

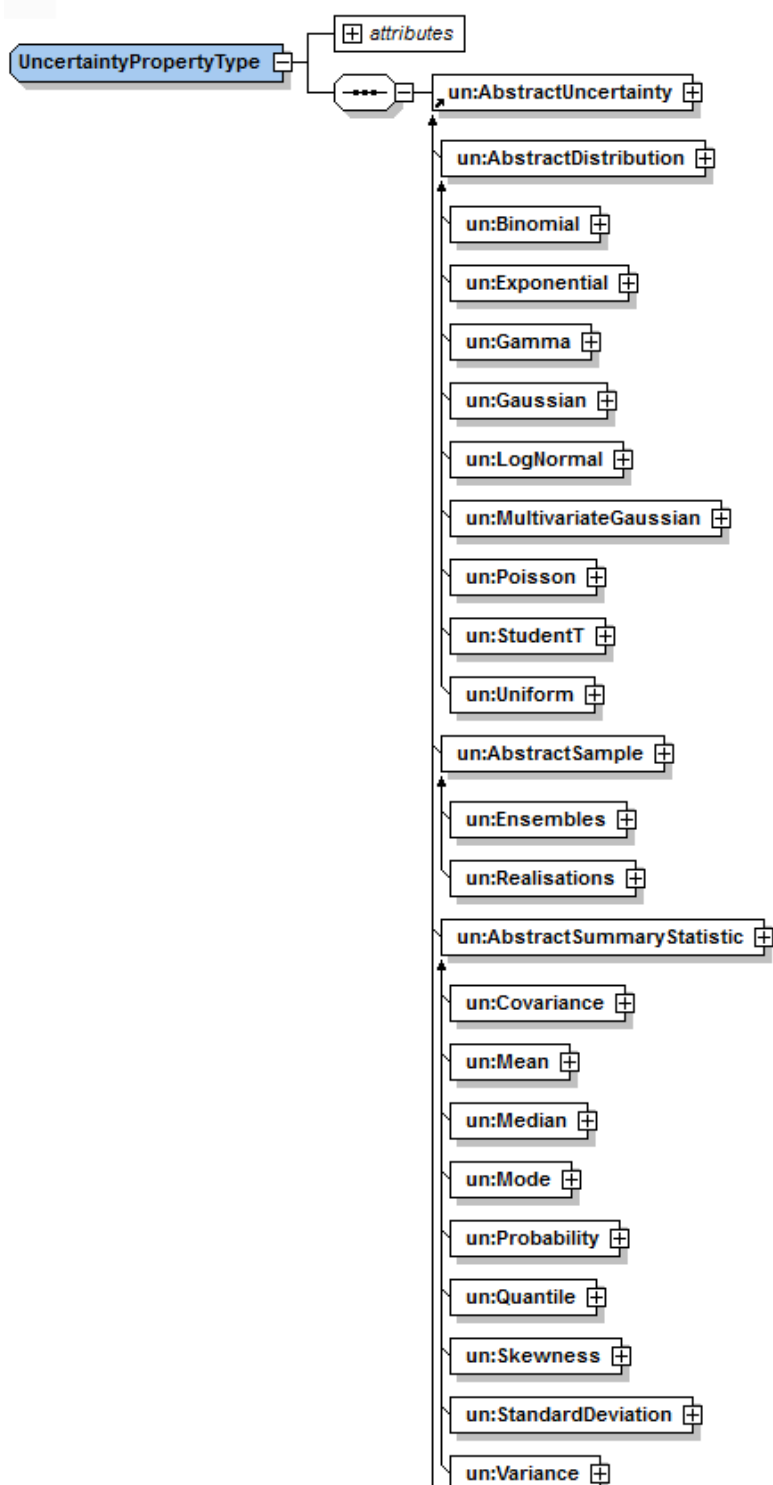


Figure 3: Schema diagram of GML UncertaintyPropertyType

5.1.1 ISO:19139 Data Quality Extension

The ISO:19139 Data Quality Extension defines a new uncertainty result type which can be used in the data quality result elements of the ISO:19139 schema files. Figure 4 shows the schema diagram of the new `DQ_UncertaintyResult_Type`. It contains two elements: A `gmd:valueUnit` element containing the unit of measure and the `gmd:value` which carries an UncertML uncertainty element.

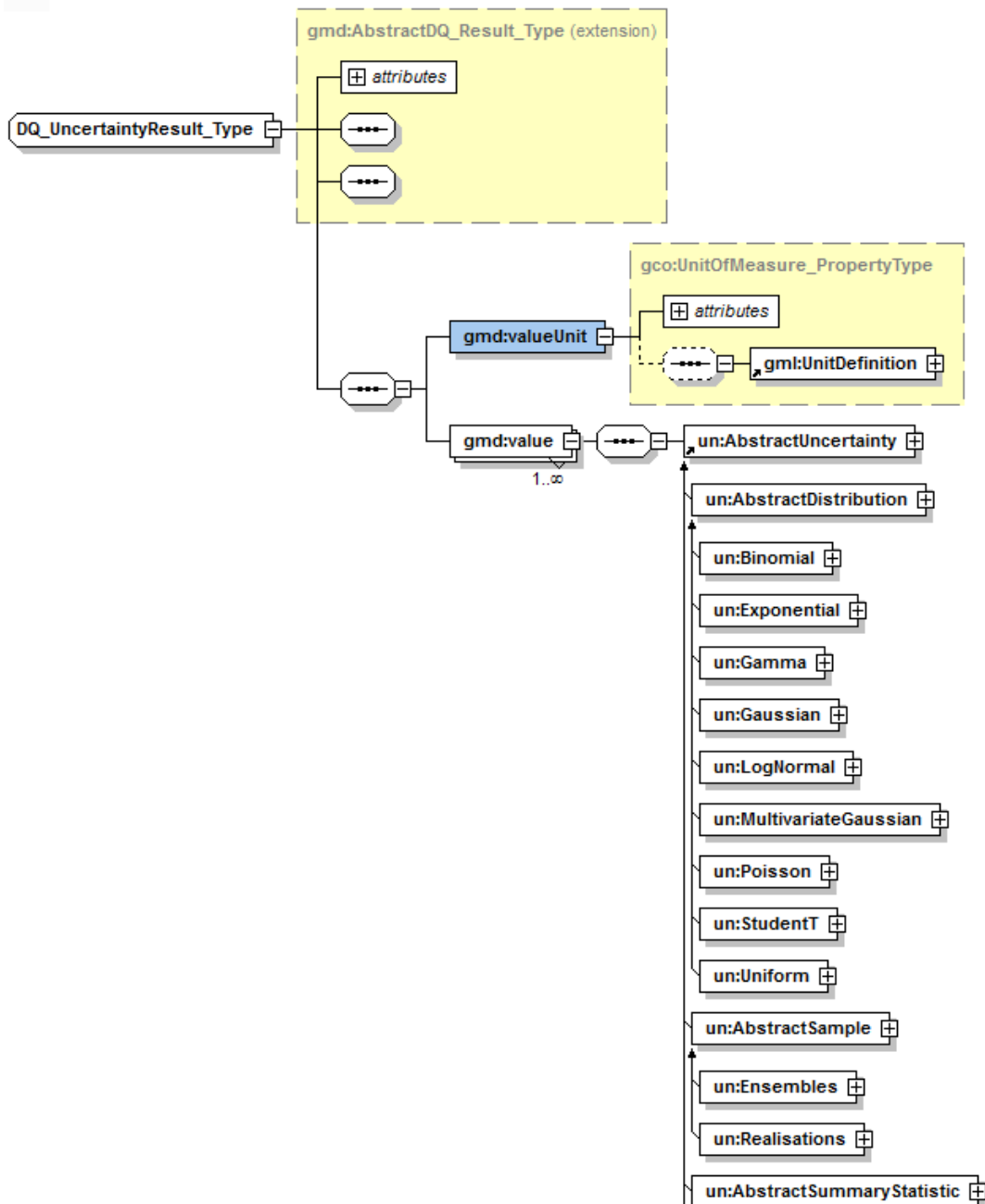


Figure 4: Schema diagram of DQ_UncertaintyResult_Type

5.2 O&M 2.0 Profile

The O&M 2.0 Profile restricts the spatial, temporal and result types of an O&M observation. It uses the elements from the GML profile defined above. The schema can be found at `Profiles/O&M/UncertWeb_OM.xsd`. Example instances which are conformant to this observation profile can be found at `Profiles/O&M/examples`.

5.2.1 Temporal restrictions

An O&M observation contains three time properties: `PhenomenonTime` depicts the time when the observation has been taken. `ResultTime` represents the time when the result value was produced. These might differ for example in case of a soil sample: the sample has been taken during `phenomenonTime` and then evaluated later in a laboratory during the `resultTime`. The `validTime` represents the time when an observation is valid. In general, there are several possibilities for encoding the temporal information. In this profile,

the temporal information can either be encoded as `gml:TimeInstant` or `gml:TimePeriod` as shown in Figure 5.

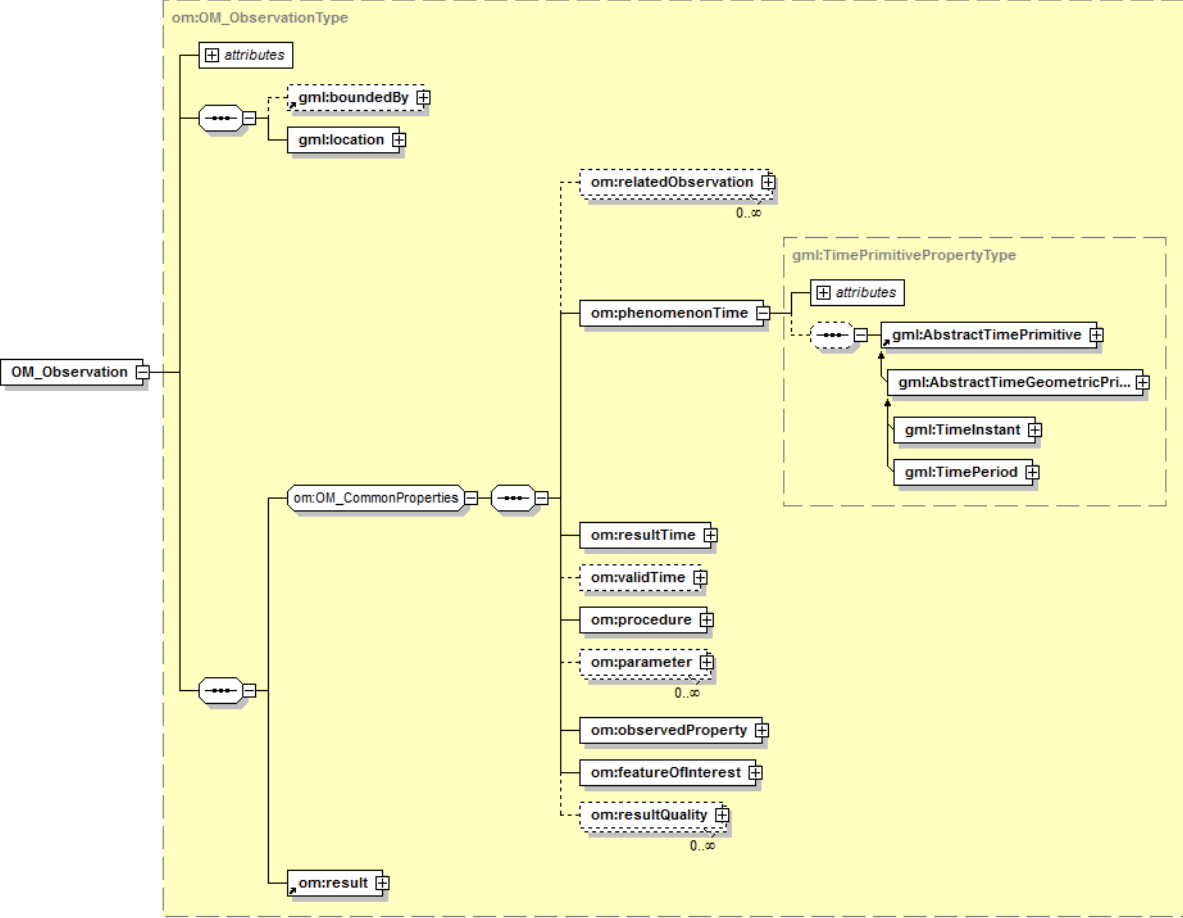


Figure 5: Schema diagram of phenomenonTime property

5.2.2 Spatial restrictions

In the basic observation model, spatial information is carried in the `featureOfInterest` property can be any feature (even non-geographic features resulting in non-spatial observations). To restrict the observations to spatial observations and to provide a common property for the sampling location of observations, the `featureOfInterest` element is restricted to `sams:SF_SpatialSamplingFeature` as shown in Figure 6.

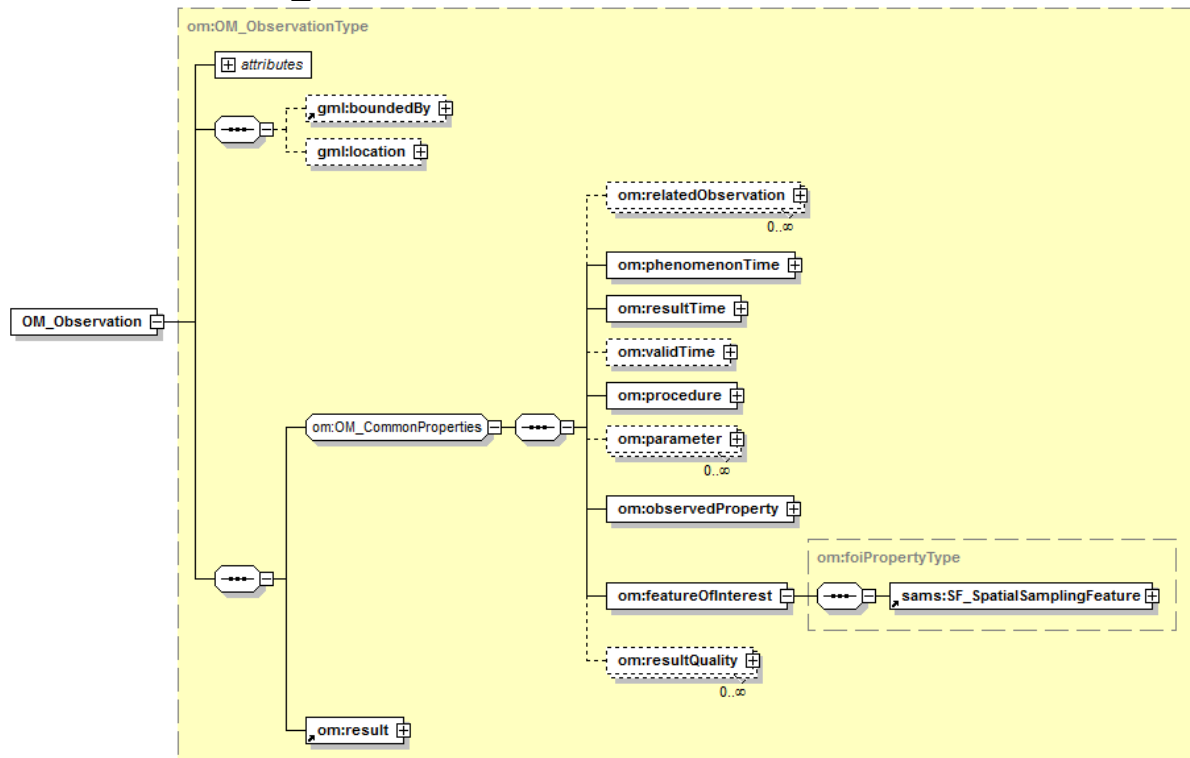


Figure 6: Schema diagram of an O&M observation with restriction of featureOfInterest to SF_SpatialSamplingFeature

We have additionally simplified the SamplingFeatures. The restrictions are defined in the schema files `Profiles/sampling/samplingFeature.xsd` and `Profiles/sampling/spatialSamplingFeature.xsd`. Figure 7 shows the schema diagram of the Spatial Sampling Features profile. All optional properties have been removed. The only mandatory properties are `sam:sampledFeature` which shall contain a reference to a real world object an observation belongs to (e.g. city or street) and `sams:shape` which has to carry the geometry of the sampling feature and thus of the observation. As can be seen in the figure, the geometries are also restricted to the ones defined in our GML profile.

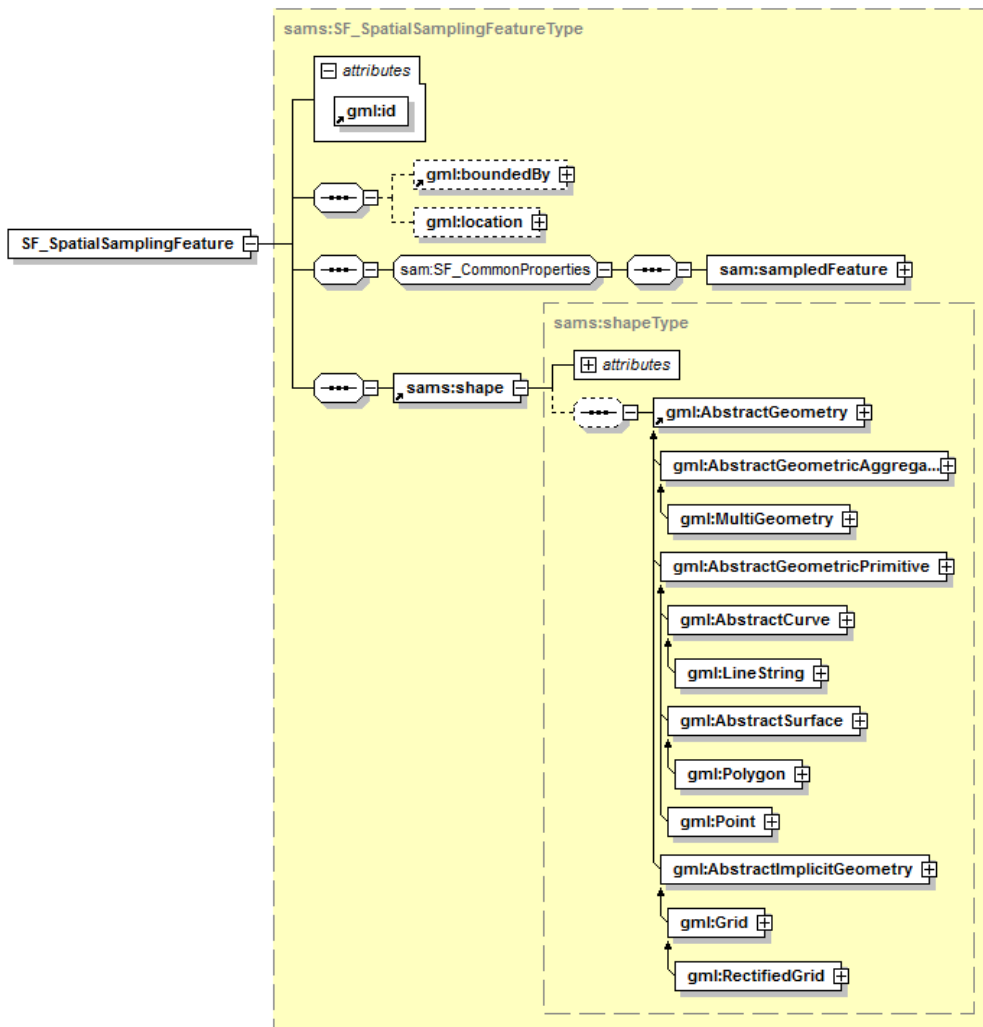


Figure 7: Schema diagram of SpatialSamplingFeature profile

5.2.3 ResultQuality restrictions

In the O&M, the resultQuality is restricted to be an ISO:19139 data quality element. Therefore, we have created the an uncertainty result type as extension of the data quality result type (see Section ISO:19139 Data Quality Extension). This enables the use of common ISO data quality elements with results carrying quantified uncertainty information. Thus, uncertainty information can be given in the `om:resultQuality` element of an observation.

5.2.4 Result restrictions

The result type of the basic observation model is not restricted (`xs:anyType`) in the XML implementation. To restrict it in UncertWeb, we have introduced a choice for the result element between Boolean, Measure, Integer, Text, Uncertainty and Reference value as shown in Figure 8. A Measure contains a double value with additional information about the unit of measurement. The uncertainty element is of type `gml:UncertainPropertyType` and enables uncertain results encoded in UncertML. The reference element enables referencing results which are provided externally, e.g. a coverage on a remote server.

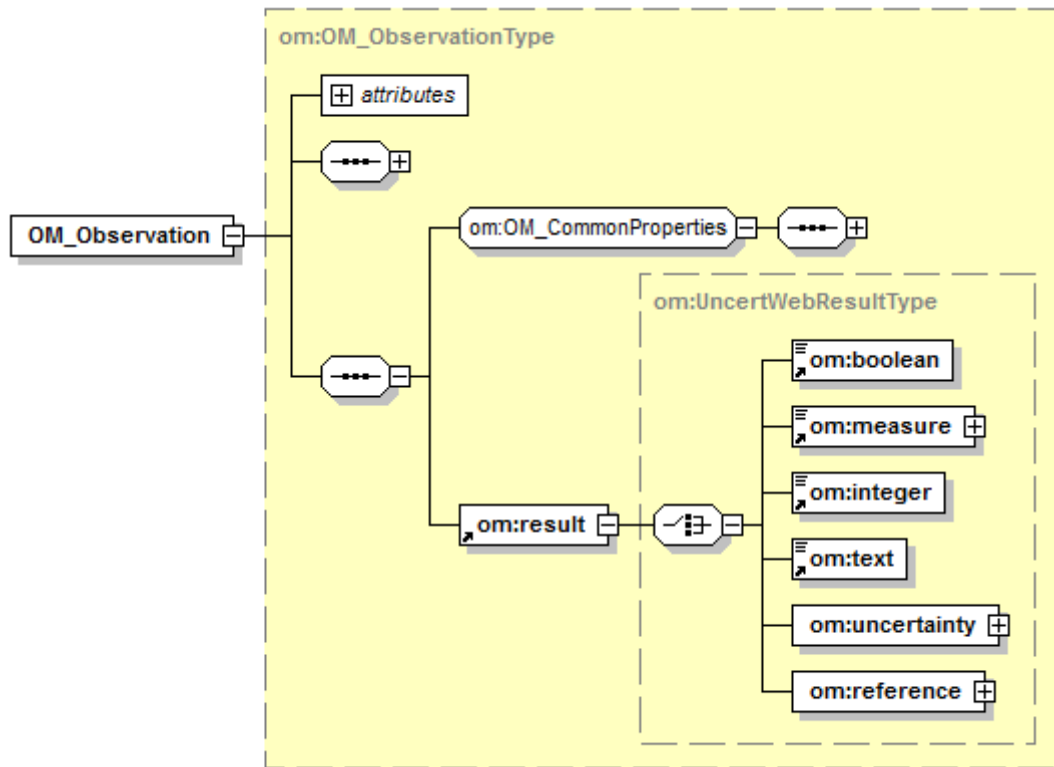


Figure 8: Schema diagram of observation result type

References

- [1] UncertWeb Project Consortium (2010). Deliverable 1.1. Technical Report, UncertWeb Project Consortium.
- [2] UncertWeb Project Consortium (2010). Deliverable 4.1, 5.1, 6.1. Technical Report, UncertWeb Project Consortium.
- [3] UncertWeb Project Consortium (2010). Deliverable 7.1. Technical Report, UncertWeb Project Consortium.
- [4] S. Cox (2007). Observations and Measurements - Part 1 - Observation schema. Open Geospatial Consortium Implementation Specification. (OGC 07-022r1)
- [5] J. Herring (2009). The Specification Model - A standard for Modular Specifications. Open Geospatial Consortium Technical Report. (OGC 08-131r3)
- [6] C. Portele (2007) OpenGIS Geography Markup Language (GML) Encoding Standard. Open Geospatial Consortium Implementation Specification. (OGC 07-036)
- [7] P. A. Vretanos (2006). Geography Markup Language (GML) - Simple Features Profile, Open Geospatial Consortium Implementation Specification. (OGC 06-049r1)