ECP-2005-GEO-038214

eWater

Report on WP8: Multi-lingual digital dictionary of hydrogeological and user-interface terminology

Deliverable number	*D-8.1 Multilingual hydrogeological vocabulary database application to provide the translation services of the eWater applications (desktop and mobile)*
Dissemination level	*Public*
Delivery date	*February 2008*
Status	*Draft/Final*
Author(s)	*Dana Čápová, Jan Sedláček, Jan Kaigl, Jiří Adamovič (Geofond, CZ)*
Other contributors:	*BRGM (FR), GBA (AU), GeoZS (ZS), GEUS (DK), LGT (LT), MAFI, (HU), SGUDS (SL), SGSS (IT), SGU (SW), TNO (NL).*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

1 Table of contents

2 Executive summary

The main objective of the eWater project is to increase cross-border accessibility and dissemination of spatial data concerning the quality, location and use of subsurface waters. Market research has shown that the demand for groundwater (hydrogeological) data is second only to demand for data on rock composition (lithology).

This data is currently stored in national databases and is therefore easily accessible only to a user fluent in the local language. As a result, hydrogeological data is confined within national borders as separate, uncorrelated data sets that are not interoperable. The consequence is that much of existing hydrogeological spatial information is difficult to access for the purpose of planning water management in an international context.

The main objective is to create an Internet system that will provide cross-border **multilingual** access to groundwater spatial data sets stored in the national databases of the participating countries. eWater Web portal will serve as a common entrance and meeting point for all those involved in cross-border water management within the EC. The portal will primarily give access to groundwater-monitoring measurements, such as water level and chemical composition, as well as to digital hydrogeological maps.

The main objective of WP 8 “Multi-lingual digital dictionary for user-interfaces and data” was the development of “Multilingual hydrogeological vocabulary database application to provide the translation services of the eWater desktop and mobile applications” (deliverable 8.1).

In order to achieve this, the following practical tasks were completed:

- Collection and translation of the required terms from the national databases.
- Harmonisation of the translations obtained from the different countries.
- Designing and implementation of the database for storing the translations.
- Development of the application allowing input of the words, their automatic translation via a common English term, and export of the database vocabulary in XML and XSL formats for use by the eWater system.
- Developing the demonstrational web site, where users can access the hydrogeological dictionary online.

As a result, the Multilingual Digital Geological Dictionary was completed that enables the translation of data and user interfaces into thirteen European languages: English and the national languages of the participants. The developed dictionary and software tools will permit the completion of a multilingual eWater web portal, allowing geoscientists and the public to acquire hydrogeological data across national borders and receive it in their native languages.

3 Introduction

The main objective of Work Package 8 “Multi-lingual digital dictionary for user-interfaces and data” was the development of “Multilingual hydrogeological vocabulary database application to provide the translation services of the eWater desktop and mobile applications” (deliverable 8.1). The purpose of the application is to allow translations of coded hydrogeological terminology on the fly, as necessary for metadata and for coded measurements. The demonstration version includes English, French, German and the native languages of the countries disseminating their data via the eWater portal. The developed translation service allows translation of codes and keywords of the borehole metadata, standardized codes of the detailed hydrogeological data and user interface texts.

The work package can be divided into four main phases:

- Developing the strategy for multilingual translation of information provided by national participants and preparing a framework for the translation.
- Constructing the multi-language database to support the Multilingual Thesaurus applications, matching the languages of the national participants.
- Developing a set of database applications for data upload, automatic translation, data viewing via a web portal, export of the dictionary in XML and XSL formats.
- Testing and documentation of the application, adding new, mainly user interface-related terms.

This Work Package was launched in month 7 of the project after WP 5 and WP 7 which defined the sets of data and metadata to be served via the eWater application. Work Package 8 has a strong link with WP 9 Service development and implementation, since the latter defined the user interface to be translated.

This report summarizes the results achieved by month 18 of the project. The following milestones defined in the project were reached at that moment:
Month 9: The framework and the strategy for the translation process were described
Month 15: The majority of applications required by this stage of the project were developed
Month 18: The compilation of the multi-language vocabulary database was completed, the translation service is ready, and the system with its applications will be brought into operation shortly.

Further activities of Work Package 8 will mainly focus on the translation of the eWater system user interfaces and user manuals, which are to be completed within Work Packages 9, 10 and 11.

4 Strategy for implementation of the Multilingual Thesaurus

The Multilingual Thesaurus (MT) is a multidisciplinary data warehouse that incorporates two disciplines: geology and hydrogeology. It has not been designed primarily as an explanatory dictionary, but as an archive of terms organized into branches.

Physically, the Multilingual Thesaurus is an Oracle relational database supplied with the additional support of data export and presentation.

The overall design of the MT was intended to be fully compatible with the Master Multilingual Thesaurus compiled during the previous European project eEarth. The resulting cross-boundary combination of European geological and hydrogeological databases now provides access to the full range of data stored with the participating geological surveys and adds value to the eWater project.

4.1 Multilingual Thesaurus implementation

- Master Multilingual Thesaurus (MMT) is a database located in Geofond (CZ). It is maintained by Geofond, filled and updated by each participant (data provider) using the applications developed
- Distributed Multilingual Thesaurus (DMT) is an export from MMT (in XML or XSL format), targeting the national participant and containing only those terms that are inside the participant's database. These terms are translated into all languages.
- A special format exported from MMT is the Portal Multilingual Thesaurus (PMT), which contains one common branch (metadata) translated to all languages (in XML format)

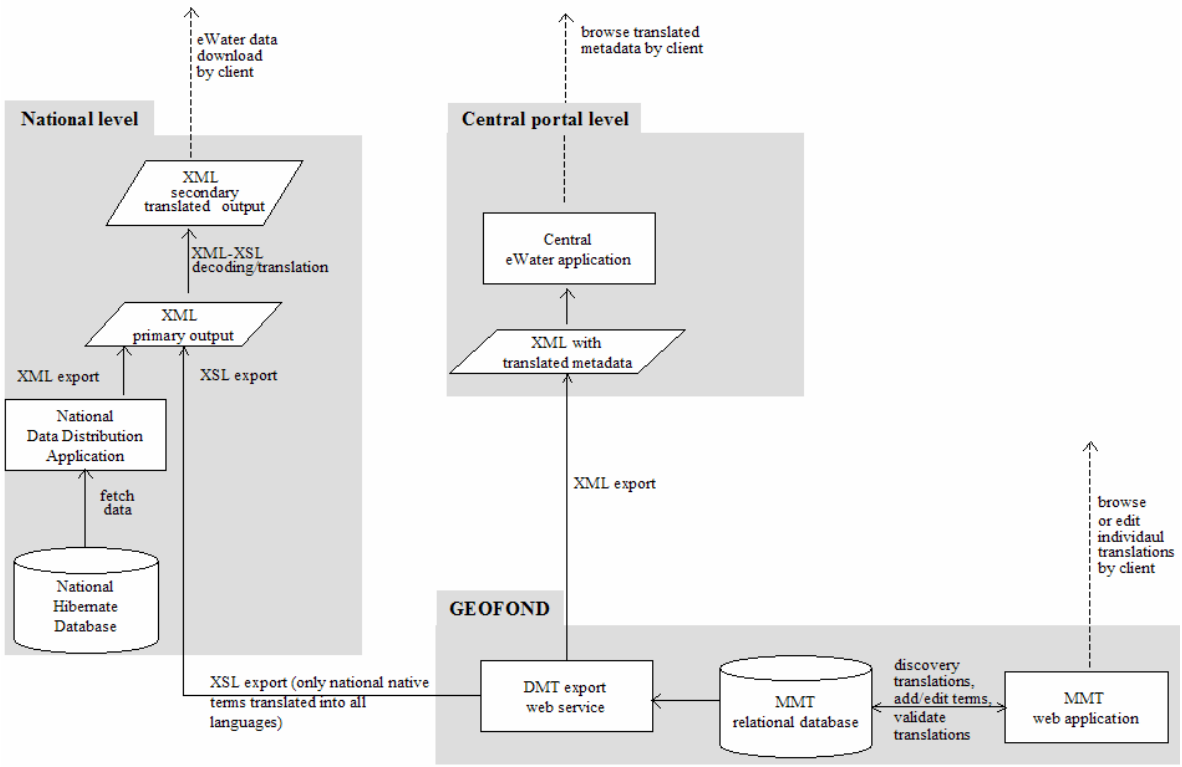
The MMT is a specifically developed application or procedural software administered with the purpose of:

- initiating the database
- primary filling with terms from national participants
- secondary translation into languages of all participants
- checking the consistency and integrity of data

The MMT is published using the web applications designed with the purpose of:

- export of relevant data to each national participant in the form of a Distributed Multilingual Thesaurus (DMT)
- export of terms for translation of metadata to the main portal level
- for internet users to browse terms and carry out translations

4.2 Multilinguality architecture



4.3 Initiation of Multilingual Thesaurus

At the first stage, initiation entails loading of the data. The second stage is concerned with the maintenance of the MT, together with addition of small numbers of terms and possible modifications in their definitions.

5 Collection of terms from partner organizations, compilation and harmonization

The work of translation within WP8 was performed at several levels in order to reflect all aspects of the multilingual data exchange between the European countries involved in the project. Hence, the following levels of the translation process were defined and carried out in an agreed order:

Basic terms: The most frequently used terms related to hydrogeology were identified and given appropriate definitions. This task was undertaken at the early stage of the project with the aim of establishing a common basis collection of data from the individual national databases.

Structured data: Collection and translation of data from the individual national databases was the principal task of the project. This required the closest co-operation in the creation of appropriate common headings, and in compilation as well as in translation of data.

Structured data consist of headings, metadata and terms.

User Interface: In the later stage of the project, user interface texts were collected from the relevant applications for translation.

5.1 Basic terms

Essential terms used to describe groundwater bodies, the groundwater table and its fluctuations, wells and well tests were collected from the national participants. Each term was annotated with an English explanation/definition. The terms & definitions were agreed by consultation with national expert groups from the participating organizations. Following agreement, they were translated into all national languages.

Aquifer; Hydraulic head *incl.* Potentiometric or Piezometric Head and Total head; Hydrogeology; Piezometer; Piezometric level; Water body; Well (water well); Spring; Borehole; Log (well log); Well test (Hydrodynamic test, Well production test); Discharge; Specific discharge; Specific yield; Drawdown; Transmissivity; Sustainable yield (safe yield); Permeability; Aquitard (aquiclude); Confined ground water; Unconfined ground water; Potential yield (Well capacity); Recovery test; Aquifer test; Pumping test; Recharge test; Unsaturated zone

Definitions were formulated after a careful review of standards published worldwide. The multiple reference literature mostly consisted of web-based glossaries, although priority was given to the universally accepted UNESCO International Glossary of Hydrogeology, see below. Other sources are listed in order of significance as used for this purpose:

(1) http://www.cig.ensmp.fr/~hydro/BIB/UNESCO/TDH000.pdf, Intl. Glossary of Hydrogeology, UNESCO 1978
(2) Domenico, P.A., Schwartz, F.W. Physical and Chemical Hydrogeology, 2nd Ed., John Wiley and Sons, New York. 1998
(3) Jackson, J.A. (1997): Glossary of Geology. 4th Ed., Amer. Geol. Inst., Alexandria, Virginia, USA. 769 pp.
(4) http://www.groundwater.org/gi/gwglossary.html, The Groundwater Foundation
(5) http://water.nv.gov/water%20planning/dict-1/ww-index.cfm, Nevada Division of Water Planning (Gary A. Horton)
(6) http://www.if.uidaho.edu/~johnson/ifiwrri/sr3/gloss.html

(7) http://www.brown.edu/Courses/GE0158/web2_revised/webglossary/hdef/hydrogeology.html

5.2 Structured data

The translation of structured data was an enormous task. It was undertaken with the purpose of establishing the Multilingual Thesaurus (MT). The MT incorporates all relevant terms from the separate national databases and also links together equivalent terms from different national databases. Because translations of highly technical terms from one language into all other languages cannot be done directly, the decision was taken in the early stages of the project that all terms should first be translated into English. *Via* English, the terms were translated in the other languages. Thus, the English terms were given the status of **master terms** in the MT. Consequently, for each participant it was necessary to translate all relevant terms from their national database into English, and all terms from other national databases into their own language *via* English. The MT then contains thirteen complete sets of terms in:
English
Czech (Geofond CZ)
Danish (Geus DK)
Dutch (TNO NL)
French (BRGM FR)
German/Austrian (GBA AT)
Hungarian (MAFI HU)
Italian (SGSS IT)
Lithuanian (LGT LT)
Slovak (GSSR SK)
Slovenian (GeoZS SI)
Spanish (IGME ES)
Swedish (SGU SV)

All terms entering the MT carry information on their source (national database), their alphanumerical codes in the individual national databases, in the cases where such codes exist, and also on the category to which they belong.

All relevant terms from the separate national databases were collected and presented in four groups:

Headings
Metadata
Terms
User Interface Terms

Headings were organized into a unified structure for the whole Consortium. This structure covers different attributes for the following types of objects:

- Well
- Well test
- Spring
- Borehole
- Point log
- Interval log

- Piezometer
- Screen
- Time series
- Time period
- Sample

The file distributed for translation contained 148 headings and their definitions.

chapter XML	XML name	headings EN	explanation EN	headings CS	explanation CS	headings DK	explanation DK
Well		Well	An artificial excavation, generally cylindrical in form and often walled in	vrt, studna	Uměle vyhloubený otvor v zemi, zpravidla válcového tvaru a s vyztuženými stěnami	Brønd	En naturlig, gravet eller boret hul i jorden, beregnet til opsamling og udnyttelse af vand
Well	name	Name of the well	Unique object name (ID)	název objektu	Jednoznačně zadané označení vrtu (identifikační číslo)	Boringsnavn	Entydig identifikation af brønden
Well	databaseName	Name of the original database	Identification of the resource database	název zdrojové databáze	Identifikace zdrojové databáze	Databasenavn	Navnet på den database som boringen kommer fra
Well	natPortalURI	URI of the National Portal	A link (Unified Resource Locator) to the National Portal	odkaz na národní portál	Odkaz na národní portál prostřednictvím jednotného zdrojového klíče	Link til national portal	Link til national hjemmeside
Well	location	Location of the well	Horizontal location of the well	souřadnice objektu	Lokalizace vrtu v horizontální rovině	Boringens placering	Boringens placering i horisontalt plan
Well	srsName	Spatial Coordinate Reference System	Spatial Coordinate Reference System according to EPGS	souřadný systém	Prostorový souřadný referenční systém podle EPGS	Koordiantsystem	Koordinatsystem efter EPGS standarden
Well	x	X-Coordinate	X-Coordinate in degrees in decimal format	souřadnice X	Souřadnice X ve stupních v desetinném tvaru	X koordinat	X koordinaten i decimalgrader
Well	y	Y-Coordinate	Y-Coordinate in degrees in decimal format	souřadnice Y	Souřadnice Y ve stupních v desetinném tvaru	Y koordinat	Y koordinaten i decimalgrader
Well	locationMethod	Method to establish the location	Method of establishing the coordinates of well location	způsob zaměření XY	Metoda použitá pro stanovení souřadnic místa vrtu	Lokaliseringsmetode	Den metode der er brugt til lokaliseringen
Well	depthFinal	Drilled depth of the well	Total length along the borehole from the earth surface, positive in metres	hloubka objektu	Celková směrná délka vrtu měřená od zemského povrchu, kladná hodnota v metrech	Boringsdybde	Den totale dybde af boringen målt fra terræn

Metadata were also organized into a unified structure for the whole Consortium before their translation. The file distributed for translation contained 118 headings.

iso	INSPIRE	tag	type	code	label_text	CZE-translation	DAN-translation	FRE-translation	DUT-translation
360	A.2.1	title	el		Resource title	název	titel	Titre	Titel
351	A.2.2.1	extent	el		Temporal extent	časový rozsah	omfang	Extension	Tijdsbestek
		timePeriod	el		Time period	Časový úsek	tids periode	Période	Periode
		from			begin	od	fra	début	Begin
		to			end	do	til	fin	Eind
		timeInstant	el		Time instant	okamžik	tidspunkt	Instant	Tijdstip
394	A.2.2.2	date	el		Date	kalendářní datum	dato	Date	Datum
395		dataType	el		data type	typ data	data type	Type de date	datatype
	A.2.2.3		co	publication	publication	zveřejnění	publikation	publication	publicatie
	A.2.2.4		co	revision	revision	revize	version	révision	revisie
	A.2.2.5		co	creation	creation	vytvoření	oprettet	création	gemaakt
343	A.2.3.1	EX_GeographicBoundingBox	el		Geographic extent of the resource - EX_GeographicBoundingBox	Geografický rozsah - ohraničující obdélník	geografisk	Extension géographique de la ressource	Geografische dekking
344		westBoundLongitude	el		west Bound Longitude	západ	grader vest	Longitude ouest	lengtegraad westkant
345		eastBoundLongitude	el		east Bound Longitude	východ	grader øst	Longitude est	lengtegraad oostkant
346		southBoundLatitude	el		south Bound Latitude	jih	grader syd	Latitude sud	breedte zuidkant
347		northBoundLatitude	el		north Bound Latitude	sever	grader nord	Latitude nord	breedte noordkant
349	A.2.3.2	geographicIdentifier	el		Geographic extent of the resource - EX_GeographicDescription	Geografický rozsah - popis geografického rozsahu	geografisk omfang	description géographique de la ressource	Geografische dekking
	A.2.4	language	el		Resource language	jazyk zdroje	sprog	langue	Taal

Terms proprietary to the individual databases were collected from each national participant in a structured form determined by the unified Headings table. For specific headings, different sets of attributes were selected to be filled in with proprietary terms (see below).

chapter XML XML name	Well	Spring	Borehole	pointLog	interval Log	Piezometer	time Series	time Period	Sample
location Method	×								
level Reference	×	×	×						
level Method	×	×	×						
Genesis	×					×			
Purpose	×	×	×						
well Type	×								
Phenomenon				×	×		×	×	×
uom				×			×	×	×
depth Method				×	×	×			
Quality Control							×	×	×

The task of translation was carried out in three steps. In **Step 1**, structured terms were collected from the national participants together with their English translations (master terms) and national codes. All collected terms were pooled, corrected for errors and sorted under their master terms. In **Step 2**, a complete table of all master terms was sent to all national participants with the request to supply translations into their own languages in case these were missing. The translations returned were critically reviewed for redundancy and inadequate translation so that during future automatic translation, unsubstantiated synonymy between terms would be avoided. The problems identified on the national level were of two different kinds: 1. the presence of two or more terms in the national database translated by the same master term (insufficiently sensitive translation); 2. the presence of two or more terms in the national database having the same national code. The attention of the respective partners was drawn to these questions and the problem was successfully solved.
Of the names of chemical substances (ca. 1,700 items), only vernacular names were selected for translation. Other substances will be displayed under their master terms in the national applications: in case of doubt, they can be unequivocally identified by their ISO codes.
The process of pooling data after Step 2, however, revealed the need for addition of several dozen new national terms in the MT. These terms, although stored in the national databases, were not submitted within Step 1 and, consequently, were not translated into other languages in Step 2. This problem was solved by **Step 3** in which a reiteration of the whole process took place. Over 2,800 items from the Terms are now kept in the MT, of which ca. 1,200 have been translated into all twelve national languages.
From Step 2 on, the whole task of translation was controlled by repeated loading of data in the MT and testing for potential errors. This approach was essential to guarantee the satisfactory function of the ultimate MT application.

5.2.1 Content of the MMT:

On 20 February 2008 the MMT contained 24 767 terms. The addition of new terms and correction of existing terms are in progress within the editing and harmonization stage.

The number of terms in different categories (in all languages):

BRANCH	NUMBER
Depth method	90
Genesis	1027
Level method	606
Level reference	257
Location method	554
Metadata	1505
Phenomenon	16519
Purpose	1246
Quality control	318
uom	1932
Well type	713

The number of terms in each of the working languages:

LANGUAGE	NUMBER
CS	1142
DA	1045
DE	2481
EN	2511
ES	1121
FR	2465
HU	1725
IT	1172
LT	1148
NL	2500
SK	2468
SL	2508
SV	2481

The number of terms in the different branches of the working languages:

BRANCH	LANG	NO	BRANCH	LANG	NO	BRANCH	LANG	NO
Depth method	FR	7	Location method	NL	34	Quality control	DA	23
Depth method	DA	7	Location method	DA	44	Quality control	EN	25
Depth method	IT	7	Location method	FR	43	Quality control	SL	28
Depth method	SV	7	Location method	DE	44	Quality control	HU	26
Depth method	SL	7	Location method	HU	42	Quality control	CS	25
Depth method	EN	7	Location method	ES	44	Quality control	FR	24
Depth method	LT	6	Location method	SK	43	Quality control	DE	25
Depth method	HU	7	Location method	IT	40	Quality control	ES	25
Depth method	NL	7	Location method	CS	44	Quality control	SK	25
Depth method	DE	7	Location method	SV	44	Quality control	SV	24
Depth method	ES	7	Location method	SL	44	Quality control	IT	16
Depth method	SK	7	Location method	EN	44	Quality control	NL	27
Depth method	CS	7	Location method	LT	44	Quality control	LT	25
Genesis	LT	80	Metadata	DA	112	Well type	SV	55
Genesis	ES	81	Metadata	HU	112	Well type	DA	53
Genesis	SV	80	Metadata	SV	114	Well type	SK	57
Genesis	NL	82	Metadata	SL	113	Well type	EN	57
Genesis	DE	80	Metadata	NL	116	Well type	ES	55
Genesis	IT	77	Metadata	IT	118	Well type	CS	57
Genesis	EN	81	Metadata	SK	116	Well type	IT	51
Genesis	HU	83	Metadata	ES	118	Well type	SL	55
Genesis	DA	63	Metadata	FR	117	Well type	LT	55
Genesis	CS	80	Metadata	EN	118	Well type	DE	55
Genesis	FR	76	Metadata	CS	116	Well type	FR	53
Genesis	SK	81	Metadata	DE	117	Well type	NL	57
Genesis	SL	83	Metadata	LT	118	Well type	HU	53
Level method	EN	47	Phenomenon	CS	498	uom	FR	154
Level method	SL	46	Phenomenon	ES	474	uom	LT	150
Level method	ES	47	Phenomenon	LT	507	uom	HU	139
Level method	DA	47	Phenomenon	DE	1838	uom	EN	153
Level method	IT	46	Phenomenon	EN	1860	uom	SV	152
Level method	SV	47	Phenomenon	IT	581	uom	ES	152
Level method	CS	47	Phenomenon	SK	1829	uom	CS	151
Level method	HU	45	Phenomenon	DA	434	uom	SK	149

BRANCH	LANG	NO	BRANCH	LANG	NO	BRANCH	LANG	NO
Level method	FR	46	Phenomenon	SV	1843	uom	NL	152
Level method	SK	46	Phenomenon	SL	1859	uom	DA	149
Level method	DE	48	Phenomenon	HU	1101	uom	IT	131
Level method	NL	48	Phenomenon	NL	1860	uom	SL	151
Level method	LT	46	Phenomenon	FR	1835	uom	DE	149
Level reference	EN	20	Purpose	SL	102			
Level reference	ES	18	Purpose	EN	99			
Level reference	LT	20	Purpose	NL	96			
Level reference	SV	20	Purpose	IT	86			
Level reference	IT	19	Purpose	DA	93			
Level reference	DE	20	Purpose	HU	98			
Level reference	NL	21	Purpose	LT	97			
Level reference	SK	20	Purpose	CS	97			
Level reference	HU	19	Purpose	FR	90			
Level reference	SL	20	Purpose	DE	98			
Level reference	CS	20	Purpose	SK	95			
Level reference	FR	20	Purpose	ES	100			
Level reference	DA	20	Purpose	SV	95			

5.3 User Interface

User Interface terms translated into the twelve national languages were taken from four different sources:

- terms for GeoNetwork operation (GIM)
- MapSphere terms (Geofond)
- terms for mobile eWater use (Geodan)
- Data Distribution Applications (GEUS)

As the process of creation of applications is still ongoing, this set of terms will probably be modified or extended depending on the parts of the applications ratified during the phase of implementation.

6 Data model for the Multilingual Thesaurus

There are two fundamental parts:

- The Master Multilingual Thesaurus (MMT) is the central database of codes from all the national participants together with their translations into all the languages of the partner organizations. This has been created by Geofond and is located on their server with Internet access.
- The Distributed Multilingual Thesaurus (DMT) is produced for specific export of translated codes. The DMT contains codes, translated values, qualifiers of translation and, in addition, an automatically generated European code (ECODE) for each of the entities incorporated in the DMT.
- The DMT is distributed to each national participant in two alternative formats
    - as a standard XML file containing all native coded values translated into each language. “Native” refers only to those terms which are included in national databases. The structure of the XML file is described below;

- as an XSL (Extensible Stylesheet) file, which contains translations of all coded values into one specified language. This file is compatible with the DataDistributionApplication, which is produced by GEUS. This xsl file is created for each language separately.

6.1 Master Multilingual Thesaurus

The centralized relational database is designed in Oracle RDBMS (Oracle 10g). The relational integrity of the objects in the database is governed internally by numerous constraints, triggers and indices. The application logic is developed partly inside the database (as stored PLSQL procedures), and partly outside by thick-client application and Internet application for thin-clients. This thin-clients application is now in the process of development.

The database must be able to support all special national diacritical characters for languages in Western, Central, Northern and Eastern Europe, therefore Unicode coding was chosen. UTF-8 support is very well incorporated into Oracle version 10g.

The current version of the MMT is in 13 languages with one master language (English), but additional languages can be added in the future.

The MMT is designed as a multidisciplinary data warehouse, covering currently two disciplines: geology and hydrogeology. All terms included are organized into branches, each branch contains similar terms describing one entity or one feature (for example “color”, “type of object” etc.)

The MMT currently contains words from databases of 15 european geological surveys.

The MMT is an open database so a new language or a new national survey can easily be added.

The MMT is designed for translation purposes only (not as an encyclopaedic reference).
The basic components of the MMT are words (not sentences). Therefore, in some languages, if a sentence were created simply by placing one word after another, it would be very likely to be grammatically incorrect.

6.1.1 Logical data model

The basic elements of the MMT data model are the keycodes and terms from the national databases. Both are organized in separate tables (see the chapter on the Physical data model). Keycodes are sorted into categories belonging to thematic branches.

The following tables contain a list of all currently presented branches with their three-letter codes.

Part Hydrogeology (subject of eWater project):

BRANCH	CODE
Heading	HDG
Conditions	NOC
Depth method	IDM

BRANCH	CODE
Genesis	ZGE
Level method	RLM
Level reference	RLR
Location method	WLO
Phenomenon	IPH
Purpose	BPU
Quality control	SQC
uom	PUO
Well type	WTY
Metadata	MET

Part Geology (subject of aready finished eEarth project):

BRANCH	CODE
Category name	CTN
Lithology	LIT
Attributes	ATR
Accuracy	ACU
Color	COL
Color intensity	COI
Elevation	ELE
Kind of water level	KWL
Object type	OBT
Well Purpose	PUR
Tests	TST
Drilling method	DRL
Coordinate system	COO
Chronostratigraphy	CRS
Drilling process	DRP

Each branch contains one category for one language; therefore the 12-language thesaurus contains twelve categories for each branch as shown in table bellow.

An example of branch and category relations

BRANCH CODE	BRANCH	CATEGORY	LANGUAGE
IPH	Phenomenon	bodový záznam - měřená veličina	CS
IPH	Phenomenon	PointLog-phenomenon	DA
IPH	Phenomenon	PointLog-phenomenon	DE
IPH	Phenomenon	parámetros del registro geofísico	ES
IPH	Phenomenon	PointLog-phenomenon	FR
IPH	Phenomenon	PointLog-phenomenon	HU
IPH	Phenomenon	parametri misurati in situ	IT
IPH	Phenomenon	tyrimai taške	LT
IPH	Phenomenon	Punt-diepte metingen	NL
IPH	Phenomenon	bodové meranie - fenomén	SK
IPH	Phenomenon	PointLog-phenomenon	SL
IPH	Phenomenon	PointLog-phenomenon	SV

This figure shows the relationship between branches and categories:

CODE	MEANING	TXCODE
51	Depth method	IDM
52	Genesis	ZGE
53	Level method	RLM
54	Level reference	RLR
55	Location method	WLO
56	Phenomenon	IPH
57	Purpose	BPU
58	Quality control	SQC
59	uom	PUO
60	Well type	WTY
61	Metadata	MET

BRANCH	CATEGORY_ID	LANGUAGE	DESCRIPTION
51	500	CS	intervalový záznam - způsob výškového zaměření
51	501	ES	método de medida de la profundidad del tramo del registro
51	502	SL	IntervalLog-depthMethod
51	503	FR	IntervalLog-depthMethod
51	504	LT	tyrimai intervale, gylio nustatymo metodas
51	505	NL	Methode bepaling diepte
51	506	DE	IntervalLog-depthMethod
51	507	SK	interval zápisu (merania) - metóda zisťovania hĺbky
51	508	DA	IntervalLog-depthMethod
51	509	EN	IntervalLog-depthMethod
51	510	HU	IntervalLog-depthMethod
51	511	SV	IntervalLog-depthMethod
51	512	IT	intervalLog profondita del log
51	513	NL	Methode bepaling diepte
51	514	ES	método de medida de la profundidad del tramo del registro
51	515	SL	Piezometer-depthMethod
51	516	FR	Piezometer-depthMethod
51	517	LT	tyrimai intervale, gylio nustatymo metodas
51	518	DE	Piezometer-depthMethod
51	519	DA	Piezometer-depthMethod
51	520	EN	Piezometer-depthMethod
51	521	HU	Piezometer-depthMethod
51	522	IT	intervalLog profondita del piezometro
51	523	CS	otevřený úsek - způsob zaměření

Interval Log

Piezometer

The MMT is designed as an open system, which permits arbitrary addition or modification of branches and languages.

Most terms are coded by national database codes, terms and codes are stored separately. The relationship between terms and their corresponding codes is shown in the following figure:

MASTER_TERM_ID	TERM_ID	LANGUAGE	DESCRIPTION
28500	28500	EN	NGF elevation
28500	28501	FR	Cote NGF
28500	34163	HU	Francia alapszint
28500	34164	IT	quota secondo il sistema di riferimento norvegese (NGF)
28500	34165	SL	NGF višina
28500	34166	DA	NGF kote
28500	34167	NL	NGF hoogte
28500	34168	SV	NGF höjd
28500	34169	DE	Normalhöhe Marseille
28500	34170	CS	nadmořská výška v systému NGF
28500	34171	ES	elevación NGF
28500	34172	SK	NGF elevácia

TERM	KEYCODE	COUNTRY	CATEGORY
28497	7	BRGM	714
28499	0	BRGM	714
28501	1	BRGM	516
28503	2	BRGM	516
28505	A	BRGM	544
28507	S	BRGM	544

Some terms are simple texts (without codes). In this case codes are automatically complemented by MT engine (these codes start with underscore), as shown in the following Figure:

MASTER_TERM_ID	TERM_ID	LANGUAGE	DESCRIPTION
28478	34058	CS	NGF
28478	34051	DA	NGF
28478	34061	DE	Normalhöhe Marseille (NGF)
28478	28478	EN	NGF
28478	34052	ES	NGF
28478	28479	FR	NGF
28478	34059	HU	Francia alapszint (NGF)
28478	34060	IT	sistema di riferimento per le quote in Norvegia
28478	34053	LT	NGF
28478	34054	NL	NGF
28478	34055	SK	NGF
28478	34056	SL	NGF
28478	34057	SV	NGF

TERM	KEYCODE	COUNTRY	CATEGORY
28473	2	BRGM	576
28475	1	BRGM	576
28477	0	BRGM	576
28479	_401	BRGM	610
28481	3	BRGM	714
28483	8	BRGM	714
28485	2	BRGM	714

The translation function of the MMT has been designed using relationships between terms. One language was selected as the master for translation (English). The Master Term, which is the term translated into the master language, is the mandatory requirement for each term and is used as an international platform for almost all bilingual translations. This approach has advantages as well as some disadvantages.

Advantage: The English term is usually understood by all participants and is relatively easy to translate into other languages.

Disadvantage: This method of translation is very sensitive to the precise meaning of the master term. For example, if the English master translation is not precise, the translation may be incorrect or ambiguous. For this reason, the "sensitivity of the master term" is defined and this has critical implications for the correctness of translations.

Master terms with low sensitivity can lead to multiple translation-synonyms. In contrast, automatically generated translations of master terms with high sensitivity are not possible. These must be translated manually.

As shown in the following Figure, each term knows its master English term (bilingual translations). The red arrows from TERM_ID=52052 point to MASTER_TERM_ID=51597, this is valid for other terms in the following Figure, where all twelve terms have the same English "Master translation":

MASTER_TERM_ID	TERM_ID	LANGUAGE	DESCRIPTION	DESC_HTML	BRANCH	SOURCE
51597	52052	CS	jardy	jardy	61	GF
51597	51937	DA	yard	yard	61	GEUS
51597	51714	DE	Yard	Yard	61	GBA
51597	51597	EN	yard	yard	61	
51597	52284	ES	yarda	yarda	61	IGME
51597	51598	FR	yard	yard	61	BRGM
51597	52512	HU	yard	yard	61	MAFI
51597	52629	IT	yard	yard	61	SGSS
51597	52401	LT	jardas	jardas	61	LGT
51597	52857	NL	yard	yard	61	TNO
51597	52167	SK	jard	jard	61	GSSR
51597	51826	SL	jard	jard	61	GEOZS

Another qualifier of the translation is an "Explicit translation", made separately by each participating organization.

The following Figure shows the Czech term "Maximální koeficient transmisivity – ČZ" translated into all twelve languages *via* the master term. This translation was made individually and has the qualifier (RELIABILITY) evaluated as "1" (yellow ellipse).
Each translation, such as the Dutch translation "maximaal doorlaatvermogen – PT" (see red rectangle) has a corresponding equivalent in all other languages. This relationship is made automatically by the MT engine, and is evaluated as RELIABILITY=3 (orange ellipse), which means a less reliable qualifier. For example, the Dutch translation "maximaal doorlaatvermogen – PT" corresponds to the Spanish translation "coeficiente de transmisividad máximo". The precision of this translation was not validated by the user and therefore may not be exact.

LANG	DESC	LANGUAGE	DESCRIPTION	RELIABILITY
CS	Maximální koeficient transmisivity - ČZ	EN	maximum coefficient of transmissivity - PT	2
CS	Maximální koeficient transmisivity - ČZ	CS	Maximální koeficient transmisivity - ČZ	9
CS	Maximální koeficient transmisivity - ČZ	HU	maximális transzmisszivitás - PT	1
CS	Maximální koeficient transmisivity - ČZ	FR	coefficient de transmissivité maximum - PT	1
CS	Maximální koeficient transmisivity - ČZ	SL	najvišja vrednost transmisivnosti - črpalni poskus	1
CS	Maximální koeficient transmisivity - ČZ	IT	coefficiente massimo di trasmissivita	1
CS	Maximální koeficient transmisivity - ČZ	ES	coeficiente de transmisividad máximo	1
CS	Maximální koeficient transmisivity - ČZ	DA	maksimal transmissivitet - PT	1
CS	Maximální koeficient transmisivity - ČZ	SV	max transmissivitet - PT	1
CS	Maximální koeficient transmisivity - ČZ	NL	maximaal doorlaatvermogen - PT	1
CS	Maximální koeficient transmisivity - ČZ	DE	maximaler Transmissivitätskoeffizient - PT	1
CS	Maximální koeficient transmisivity - ČZ	LT	vandens pralaidumo koeficientas - IŠP.	1
CS	Maximální koeficient transmisivity - ČZ	SK	maximálny koeficient prietočnosti - ČS	1
CS	Maximální koeficient transmisivity - ČZ	NL	maximum doorlaatvermogen - PT	1

1 = explicit translation

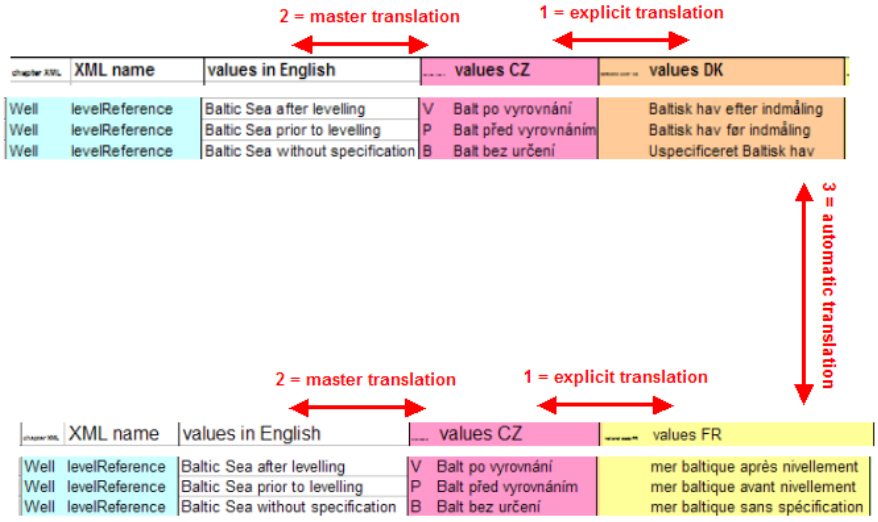
LANG	DESC	LANGUAGE	DESCRIPTION	RELIABILITY
NL	maximum doorlaatvermogen - PT	EN	maximum coefficient of transmissivity - PT	2
NL	maximum doorlaatvermogen - PT	CS	Maximální koeficient transmisivity - ČZ	1
NL	maximum doorlaatvermogen - PT	HU	maximális transzmisszivitás - PT	3
NL	maximum doorlaatvermogen - PT	FR	coefficient de transmissivité maximum - PT	3
NL	maximum doorlaatvermogen - PT	SL	najvišja vrednost transmisivnosti - črpalni poskus	3
NL	maximum doorlaatvermogen - PT	IT	coefficiente massimo di trasmissivita	3
NL	maximum doorlaatvermogen - PT	ES	coeficiente de transmisividad máximo	3
NL	maximum doorlaatvermogen - PT	DA	maksimal transmissivitet - PT	3
NL	maximum doorlaatvermogen - PT	SV	max transmissivitet - PT	3
NL	maximum doorlaatvermogen - PT	NL	maximaal doorlaatvermogen - PT	13
NL	maximum doorlaatvermogen - PT	DE	maximaler Transmissivitätskoeffizient - PT	3
NL	maximum doorlaatvermogen - PT	LT	vandens pralaidumo koeficientas - IŠP.	3
NL	maximum doorlaatvermogen - PT	SK	maximálny koeficient prietočnosti - ČS	3
NL	maximum doorlaatvermogen - PT	NL	maximum doorlaatvermogen - PT	9

13 = synonym

3 = automatic translation

Principle of evaluation of relationships between terms:

The following figure shows the translations made by the participating organizations, when GEOFOND terms were translated into Danish (upper table) and into French (lower table) *via* the English master terms. These translations were evaluated as "explicit translations" with qualifier=1.
In contrast, the relationships between Danish and French translations were created automatically by the MT engine and evaluated as "automatic translations" with qualifier=3.



Synonyms produced during the automatic process of translation are evaluated. Synonyms arise by the same process as automatic translations, the only difference being that both terms are in

the same language. As discussed above, in such a case, the "sensitivity of the master term" is also critical for the selection of synonyms. All terms in the same language that have exactly the same master term are synonyms. Terms are stored in the vector table TERMS, the relationship between IDs is stored in the table TERM_REL with a reliability qualifier 13 (a flag for automatic synonym).

The quality of translation is defined in the domain table DOM_RELIABILITY. A list of the defined qualifiers is shown in the table below.

CODE	MEANING	DESCRIPTION
1	explicit translation	term1 - term2 (user confirmed translation, different language)
2	master translation	term - master (user defined translation)
3	auto translation	term1 - master - term2 (automated translation, different language)
9	self relationship	term - term (for integrity purpose)
11	explicit synonym	term1 - term2 (user confirmed synonym, same language)
13	auto synonym	term1 - master - term2 (automated translation, same language)
99	invalid relationship	

In the process of automatic translation, codes 2, 9, and 13 are set during the bulk loading process, codes 1, 3, 9, and 13 are set during the bulk translation process, and codes 1, 11, and 99 are set manually using the internet MMT application.

The numeric qualifier "CODE" increases with decreasing quality of translation. Formal code "9" is suggested as a flag for a national term which is translated. Synonyms are given a code greater than 10. The formal code "99" is reserved for the internet MMT application, allowing the user to flag a relationship for possible deletion in future.
During the interactive user control process, all entries coded "3" should be changed to code "1" if the translation is acceptable, or to code "99" if the translation is to be rejected. All entries for synonyms which are coded "13" should be changed to "11" if the synonym is acceptable or to "99" if it is to be rejected.
Under optimum conditions, the final edition of the MMT should not contain any entries coded "3" or "13".

The translation process is carried out during bulk loading of data or on-line using the internet MMT application. This process is based on the matching of master terms. It is split into two stages:

- loading new terms with their master translation
- loading translated terms matched against the existing terms corresponding to the same English master terms

The following diagrams illustrate these two stages:

Diagram 1: process "loading new terms":

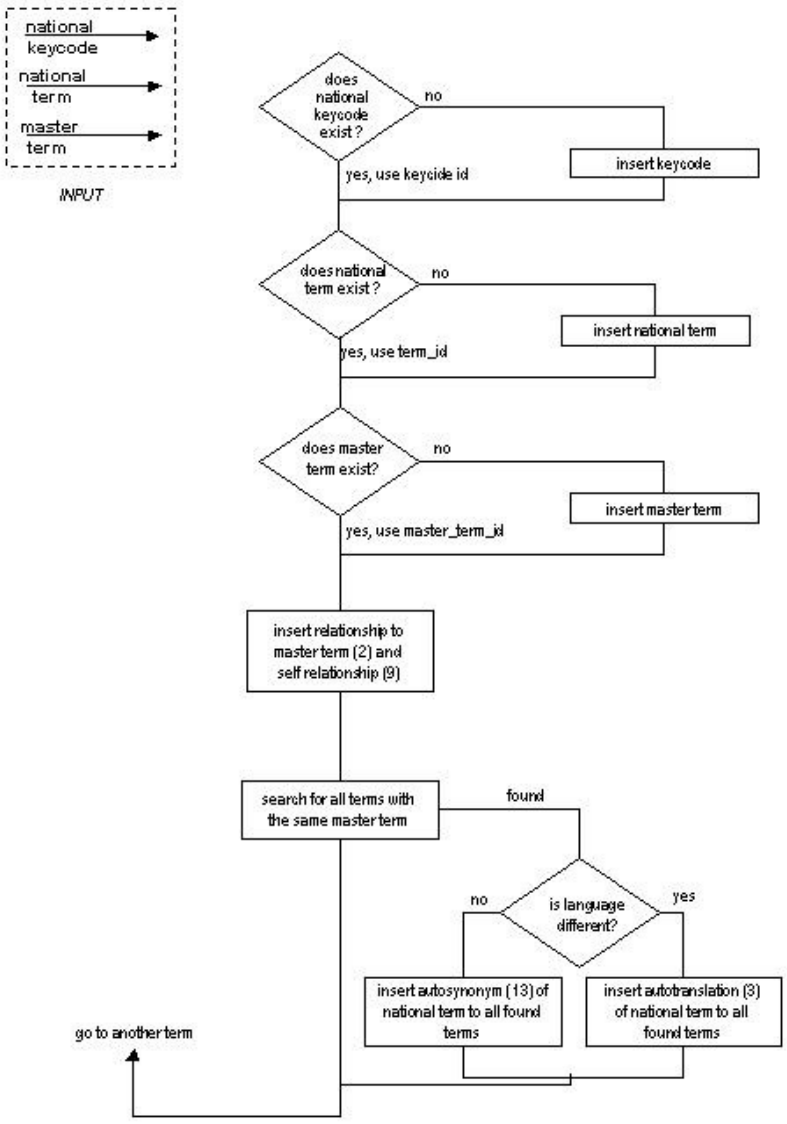
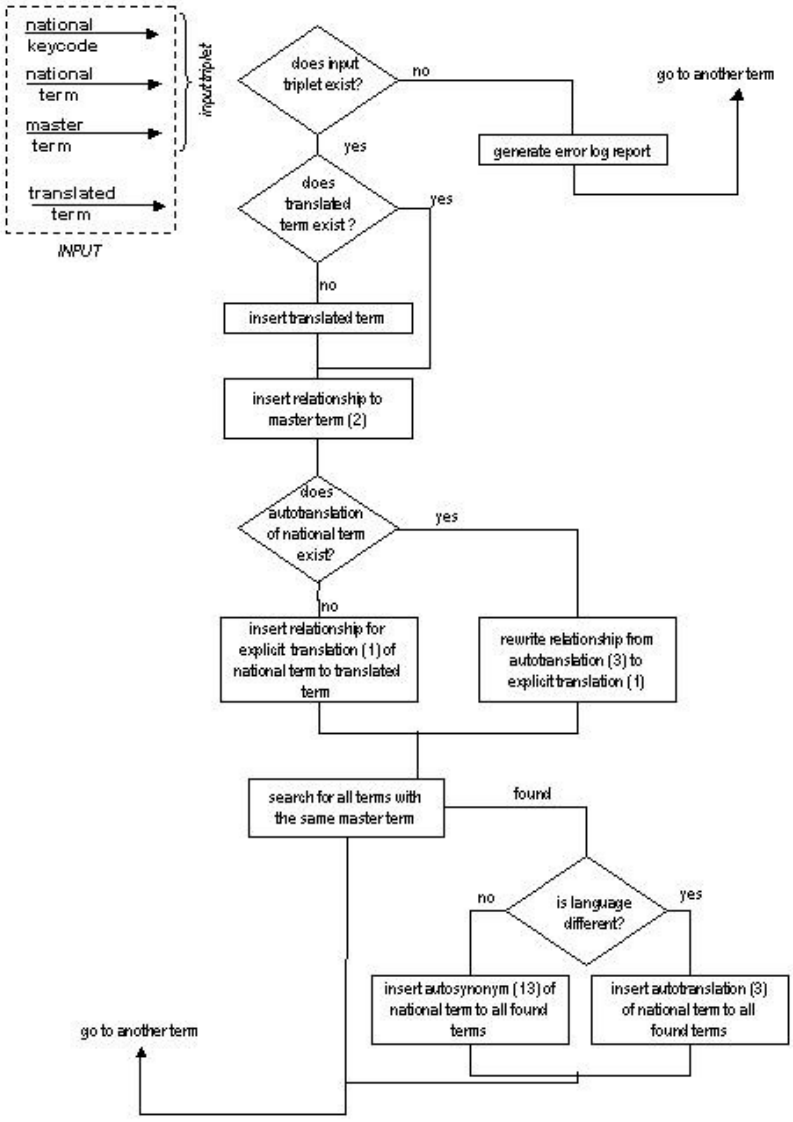


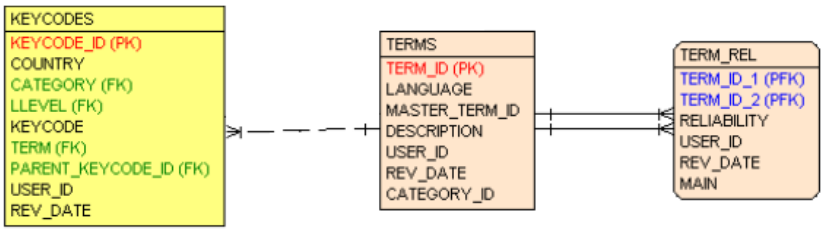
Diagram 2: process – " loading translated terms":



6.1.2 Physical data model

Codes from national databases, terms for these codes and the relationship between terms are stored in the core tables of the MMT database store.

Core tables with national codes (KEYCODE), with terms (TERMS) and terms relationships (TERM_REL):



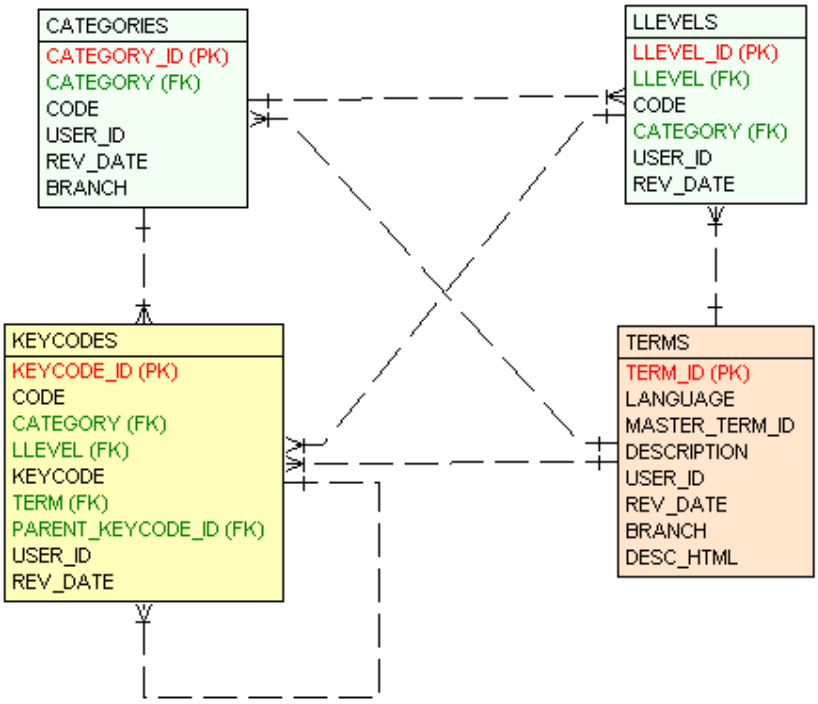
Categorization of terms and value control is incorporated in additional tables called domain tables.

National codes are organized in one vector table „KEYCODES“ with a coded value from the national database („KEYCODE“) and the relationship to the code description („TERM“), which is stored in the second vector table „TERMS“. Terms without keycodes are also possible, but in this case during the loading process a keycode is created automatically as an autonumber with an underscore prefix (for example, in the case of headings or chapter names a keycode such as “_1”, or “_2” etc. is generated automatically).

All words are contained inside the vector table „TERMS” („DESCRIPTION“), all bilingual translation couples are contained within the table of relationships between terms – „TERM_REL“ („TERM_ID1“, „TERM_ID2“). The relation between the table „TERMS“ and the table „KEYCODES“ permits a single keycode for one or several terms. The relation between the table „TERMS“ and the table „TERM_REL“ permits bidirectional relationships to be stored.

Each term must have a master translation (in English), which is stored as a regular term. The relationship between a term and the master term is stored directly in the table „TERMS“ within the column „MASTER_TERM_ID“ and also in the table „TERM_REL“ as a bidirectional relationship.

Terms are classified into categories; each category belongs to a branch. The following chart shows a part of the database structure responsible for classification of data.



Keycodes can be optionally linked to the level; this relationship is used for hierarchical ordering of keycodes used, for example, for stratigraphic purposes inside the geological part of the MMT. The level expresses, for example, the stratigraphic level such as “era”, “erathem” etc. Moreover, stratigraphic keycodes are linked to the parent stratigraphic keycode (see cyclic foreign relationship between columns “keycode_id” and “parent_keycode_id”).

Data consistency is controlled by several domain tables connected to data tables with foreign keys. These domains control valid (allowable) values inside columns. There are domains responsible for:

- list of national geological surveys (DOM_COUNTRY)
- list of languages currently incorporated (DOM_LANGUAGE)
- list of all users with read/write access to the MMT (DOM_USER)
- user roles such as “DBA”, “creator”, “editor”, “guest” (DOM_ROLES)

Data integrity is controlled by several triggers.

Referential integrity is managed by primary and foreign keys, check or null constraints and unique indices. Procedural support is ensured with packages written in PLSQL, by which the export of data and statistical calculations are undertaken.

7 Applications for DMT export and for MMT maintenance

7.1 DMT export

The Distributed Multilingual Thesaurus (DMT) is the logical expression for the special export of part of the MMT. It is targeted to individual participants and contains keycodes, terms and all available translations for that participant. For example, the DMT for GEUS contains only GEUS keycodes and terms, together with all available translations of the GEUS terms.

The DMT is currently available in two formats:

- a) universal XML export with SchemaDefinition file XSD
- b) specially tailored Extensible Stylesheet file XSL, compatible with the Data Distribution Application

7.1.1 DMT export as XML table

a) The physical structure of the DMT is a spreadsheet, in which each row contains one keycode, one term and several additional attributes belonging to this term.
The header of the XML file contains information about UTF-8 encoding, which supports special national characters (such as Lithuanian, Czech, Danish etc.) to be transferred. Moreover, an html representation of each term is added (see hterm="jiný“ in the following example).
The header of the XML file also contains the target geological survey (see tag <translation survey="GEUS"> in the following example):

<?xml version="1.0" encoding="UTF-8"?>
<translation survey="GEUS" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.ewater.eu/mmt2dmt ./ewater.mmt2dmt.xsd" xmlns="http://www.ewater.eu/mmt2dmt">
<term ecode="ZGE00059" lang="CS" qual="1" catid="575" cat="vrt - zůsob výškového zaměření" ncode="A" term="jiný"
hterm="jiný" source="" />
<term ecode="ZGE00059" lang="DA" qual="9" catid="575" cat="Borehole-levelMethod" ncode="A" term="Andet" hterm="Andet"
source="GEUS" />
<term ecode="ZGE00059" lang="DE" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="andere" hterm="andere"
source="" />
<term ecode="ZGE00059" lang="EN" qual="2" catid="575" cat="Borehole-levelMethod" ncode="A" term="other" hterm="other"
source="" />
<term ecode="ZGE00059" lang="ES" qual="1" catid="575" cat="método de nivelación del sondeo" ncode="A" term="otros" hterm="otros"
source="" />
<term ecode="ZGE00059" lang="FR" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="autre" hterm="autre"
source="" />
<term ecode="ZGE00059" lang="HU" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="egyéb" hterm="egyéb"
source="" />
<term ecode="ZGE00059" lang="IT" qual="1" catid="575" cat="metodo di rilevamento della quota del pozzo" ncode="A" term="Altro"
hterm="Altro" source="" />
<term ecode="ZGE00059" lang="LT" qual="1" catid="575" cat="gręžinio aukščio nustatymo metodas" ncode="A" term="kita"
hterm="kita" source="" />
<term ecode="ZGE00059" lang="NL" qual="1" catid="575" cat="niveaubepaling boring" ncode="A" term="anders" hterm="anders"
source="" />
<term ecode="ZGE00059" lang="SK" qual="1" catid="575" cat="výškový systém - vrty (určenie)" ncode="A" term="iný"
hterm="iný" source="" />
<term ecode="ZGE00059" lang="SL" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="ostalo" hterm="ostalo"
source="" />
<term ecode="ZGE00059" lang="SV" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="övrigt" hterm="övrigt"
source="" />

Each data row contains the following tags:

- **ecode**: automatically generated unique eWater code derived from branch (3 character prefix) and 5 digits of auto number
- **lang**: language code for the term
- **qual**: numerical qualifier for translation (such as master, explicit, automatic translation or automatic synonym)
- **catid**: category numerical identifier
- **cat**: category description translated to the corresponding language
- **ncode**: national keycode (in this example code from GEUS database)
- **term**: term description in Unicode UTF-8 coding
- **hterm**: term description in html format (diacritical characters are represented as html codes)
- **source**: optional tag used to denote the name of the source database from which the keycode comes (in this example, the Danish term is native to the GEUS database)

The XML file is formally controlled by the SchemaDefinition file (XSD file), which contains several domains responsible for correct numerical or textual values of languages, participating organizations or quality of translation. The XSD file describes the content of the XML file, the format of each tag when the value is mandatory etc. Appendix A1 contains a full listing of the XSD file used for the eWater XML Distributed Multilingual Thesaurus.
This format for the DMT export is universal but requires additional programming work during the creation of the Data Distribution Application.

The following figure shows several steps necessary for the translation, which must be done within the Data Distribution Application. The example below shows a translation of the Danish coded term „A“ („Andet“) from category „Borehole-levelMethod“ into Czech:

- step 1 - find category ID for this term

select catid from .... where lang='EN' and cat='Borehole-levelMethod' ➔ catid = 575

```
<term ecode="ZGE00059" lang="CS" qual="1" catid="575" cat="vrt - způsob výškového zaměření" ncode="A" term="jiný" hterm="jiný" source="" />
<term ecode="ZGE00059" lang="DA" qual="9" catid="575" cat="Borehole-levelMethod" ncode="A" term="Andet" hterm="Andet" source="GEUS" />
<term ecode="ZGE00059" lang="DE" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="andere" hterm="andere" source="" />
<term ecode="ZGE00059" lang="EN" qual="2" catid="575" cat="Borehole-levelMethod" ncode="A" term="other" hterm="other" source="" />
<term ecode="ZGE00059" lang="ES" qual="1" catid="575" cat="método de nivelación del sondeo" ncode="A" term="otros" hterm="otros" source="" />
<term ecode="ZGE00059" lang="FR" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="autre" hterm="autre" source="" />
<term ecode="ZGE00059" lang="HU" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="egyéb" hterm="egyéb" source="" />
<term ecode="ZGE00059" lang="IT" qual="1" catid="575" cat="metodo di rilevamento della quota del pozzo" ncode="A" term="Altro" hterm="Altro" source="" />
<term ecode="ZGE00059" lang="LT" qual="1" catid="575" cat="gręžinio aukščio nustatymo metodas" ncode="A" term="kita" hterm="kita" source="" />
<term ecode="ZGE00059" lang="NL" qual="1" catid="575" cat="niveaubepaling boring" ncode="A" term="anders" hterm="anders" source="" />
<term ecode="ZGE00059" lang="SK" qual="1" catid="575" cat="výškový systém - vrty (určenie)" ncode="A" term="iný" hterm="iný" source="" />
<term ecode="ZGE00059" lang="SL" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="ostalo" hterm="ostalo" source="" />
<term ecode="ZGE00059" lang="SV" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="övrigt" hterm="övrigt" source="" />
```

- step2 - find this term translated into Czech

select term from .... where catid = 575 and lang='CZ' and ncode 'A' ➔ term = jiný

```
<term ecode="ZGE00059" lang="CS" qual="1" catid="575" cat="vrt - způsob výškového zaměření" ncode="A" term="jiný" hterm="jiný" source="" />
<term ecode="ZGE00059" lang="DA" qual="9" catid="575" cat="Borehole-levelMethod" ncode="A" term="Andet" hterm="Andet" source="GEUS" />
<term ecode="ZGE00059" lang="DE" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="andere" hterm="andere" source="" />
<term ecode="ZGE00059" lang="EN" qual="2" catid="575" cat="Borehole-levelMethod" ncode="A" term="other" hterm="other" source="" />
<term ecode="ZGE00059" lang="ES" qual="1" catid="575" cat="método de nivelación del sondeo" ncode="A" term="otros" hterm="otros" source="" />
<term ecode="ZGE00059" lang="FR" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="autre" hterm="autre" source="" />
<term ecode="ZGE00059" lang="HU" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="egyéb" hterm="egyéb" source="" />
<term ecode="ZGE00059" lang="IT" qual="1" catid="575" cat="metodo di rilevamento della quota del pozzo" ncode="A" term="Altro" hterm="Altro" source="" />
<term ecode="ZGE00059" lang="LT" qual="1" catid="575" cat="gręžinio aukščio nustatymo metodas" ncode="A" term="kita" hterm="kita" source="" />
<term ecode="ZGE00059" lang="NL" qual="1" catid="575" cat="niveaubepaling boring" ncode="A" term="anders" hterm="anders" source="" />
<term ecode="ZGE00059" lang="SK" qual="1" catid="575" cat="výškový systém - vrty (určenie)" ncode="A" term="iný" hterm="iný" source="" />
<term ecode="ZGE00059" lang="SL" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="ostalo" hterm="ostalo" source="" />
<term ecode="ZGE00059" lang="SV" qual="1" catid="575" cat="Borehole-levelMethod" ncode="A" term="övrigt" hterm="övrigt" source="" />
```

This DMT has been distributed to all participants *via* e-mail but, at the request of a user, the dynamic export of the DMT by XML data transfer can be made using an internet MMT application.

7.1.2 DMT export as XSL file

Another special export format of DMT is Extensible Stylesheet file XSL. This file format is developed especially for decoding or translating XML tags.
The XSL file is tailored to the XML file, which is produced by the Data Distribution Application. Each tag from the DDA XML export file has its definition in the XSL file determining how to handle it.

Here is an example of a translation of the XML tags using XSL support.
The following part of the XML file contains several attributes of a well produced by the DDA application:

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<?xml-stylesheet type="text/xsl" href="well.xsl"?>

  <soapenv:Body>
    <getWellByIdOutputMsg xmlns="http://www.ewater.eu/DataDistribution">
      <ns1:Well xmlns:ns1="http://www.ewater.eu">
        <ns1:databaseName>Jupiter</ns1:databaseName>
        <ns1:finalDepth>125.0</ns1:finalDepth>
        <ns1:levelMethod>G</ns1:levelMethod>
        <ns1:location>
          <ns1:srsName>GEUS</ns1:srsName>
          <ns1:x>15.0428</ns1:x>
          <ns1:y>55.0036</ns1:y>
        </ns1:location>
        <ns1:locationMethod>I</ns1:locationMethod>
        <ns1:name>248.58</ns1:name>

        <ns1:purpose>VV</ns1:purpose>
      </ns1:Well>
    </getWellByIdOutputMsg>
  </soapenv:Body>
</soapenv:Envelope>
```

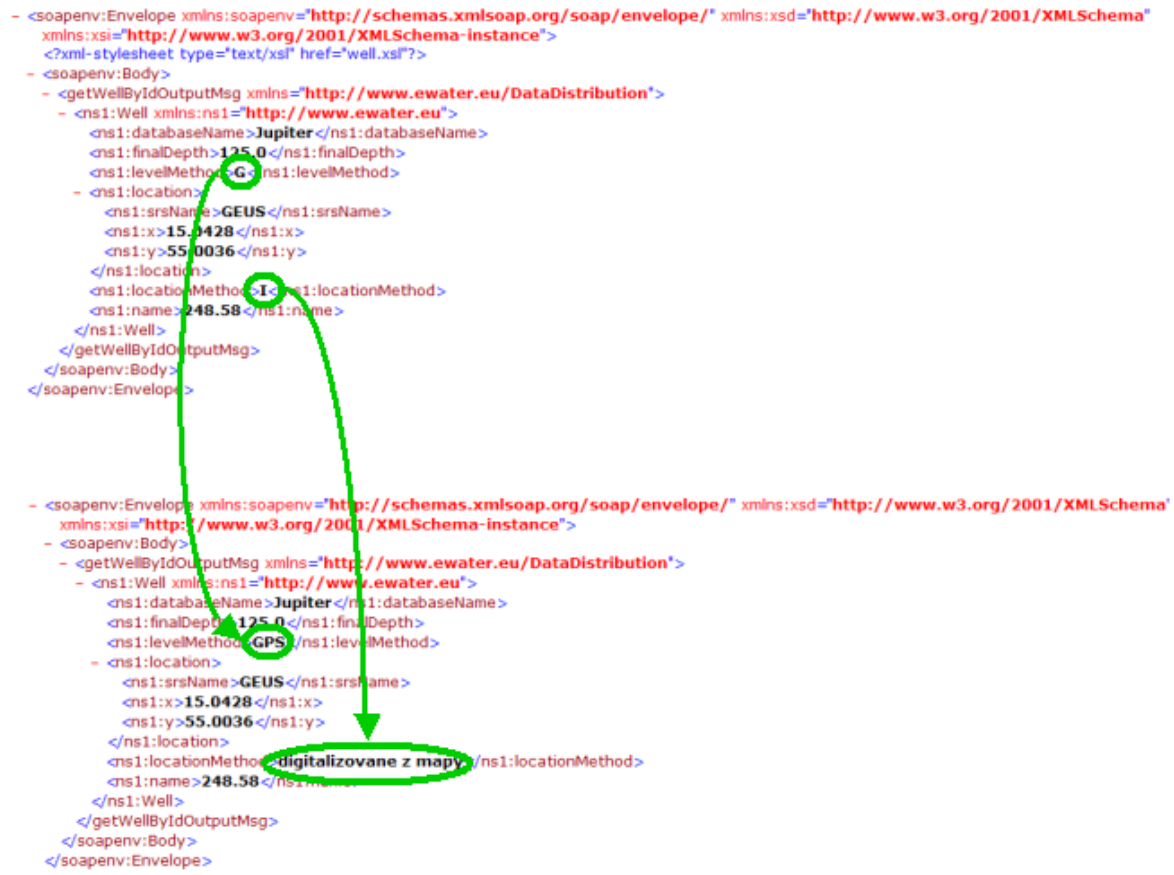
The following part of the XSL file contains a small program code defining the manipulation with all tags. Some tags, like FinalDepth in the following example, are simply rewritten. Other tags, like LocationMethod, are decoded. During this decoding process, codes are translated to the selected language. In the following example, coded values of location method (tag locationMethod) or level method (tag levelMethod) are translated into Slovak :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE xsl:stylesheet [<!ENTITY nbsp " "> ]>
```

```
<xsl:stylesheet version="1.0" xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform" xmlns:ns1="http://www.ewater.eu">
  <xsl:output method="xml" indent="yes" encoding="UTF-8" />

<xsl:template match="/">
<table>
  <xsl:apply-templates/>
</table>
</xsl:template>

  <xsl:template match="ns1:Well">
     <tr><td><xsl:text>&lt;ns1:databaseName&gt;</xsl:text>
     <xsl:value-of select="ns1:databaseName"/>
     <xsl:text>&lt;/ns1:databaseName&gt;</xsl:text></td></tr>

     <tr><td><xsl:text>&lt;ns1:finalDepth&gt;</xsl:text>
     <xsl:value-of select="ns1:finalDepth"/>
     <xsl:text>&lt;/ns1:finalDepth&gt;</xsl:text></td></tr>

     <tr><td><xsl:apply-templates select="ns1:levelMethod"/></td></tr>

     <tr><td><xsl:apply-templates select="ns1:locationMethod"/></td></tr>
  </xsl:template>

  <xsl:template match="ns1:locationMethod">
  <tr><td><xsl:choose>
    <xsl:when test=". = 'L'">&lt;ns1:locationMethod&gt;letecká snímka&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'DEC'">&lt;ns1:locationMethod&gt;Decca&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'DG'">&lt;ns1:locationMethod&gt;rozdielové GPS&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'K'">&lt;ns1:locationMethod&gt;digitálna mapa&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'T'">&lt;ns1:locationMethod&gt;digitálna mapa 1:10000&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'TK'">&lt;ns1:locationMethod&gt;digitálna technická mapa&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'D'">&lt;ns1:locationMethod&gt;digitalizované z mapy&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'O'">&lt;ns1:locationMethod&gt;vzdialenosť od rohu mapy&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'KD'">&lt;ns1:locationMethod&gt;zo zákresu, náčrtu&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'G'">&lt;ns1:locationMethod&gt;systém zisťovania miesta pomocou družíc
(GPS)&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'F'">&lt;ns1:locationMethod&gt;ortofoto&lt;/ns1:locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'A'">&lt;ns1:locationMethod&gt;iný&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'SY'">&lt;ns1:locationMethod&gt;výšková mapa&lt;/locationMethod&gt;</xsl:when>
    <xsl:when test=". = 'I'">&lt;ns1:locationMethod&gt;merač&lt;/locationMethod&gt;</xsl:when>
    <xsl:otherwise></xsl:otherwise>
  </xsl:choose></td></tr>
  </xsl:template>

  <xsl:template match="ns1:levelMethod">
  <tr><td><xsl:choose>
    <xsl:when test=". = 'P'">&lt;ns1:levelMethod&gt;differential GPS&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'H'">&lt;ns1:levelMethod&gt;digital elevation model&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'M'">&lt;ns1:levelMethod&gt;digital map&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'D'">&lt;ns1:levelMethod&gt;digital terrain model&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'E'">&lt;ns1:levelMethod&gt;echosounder&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'S'">&lt;ns1:levelMethod&gt;from a sketch&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'G'">&lt;ns1:levelMethod&gt;GPS&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'N'">&lt;ns1:levelMethod&gt;levelled elevation&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'F'">&lt;ns1:levelMethod&gt;map in the field&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'K'">&lt;ns1:levelMethod&gt;map in the office&lt;/levelMethod&gt;</xsl:when>
    <xsl:when test=". = 'A'">&lt;ns1:levelMethod&gt;other&lt;/levelMethod&gt;</xsl:when>
    <xsl:otherwise></xsl:otherwise>
  </xsl:choose></td></tr>

  </xsl:template>
</xsl:stylesheet>
```

Extensible Stylesheet Language (XSL) is primarily used for the transformation of XML documents into HTML or other XML “human-readable” documents. In our case, this file is used for converting the original XML file into the same structured XML file, but with decoded and translated original coded values. Therefore, each definition for decoding also contains opening and closing tags (<ns1:levelMethod> and </levelMethod&gt).

The following figure shows the original coded values (upper part) and corresponding decoded and translated values (lower part). Level method “G” is translated to “GPS”, location method “I” is translated into Slovak “digitalizovane z mapy”.



Summary to the XSL strategy

- The XSL file is generated for each language and each participating organization separately (total number for 12 surveys and 12 languages is 144 xsl files) . Each implementation of the Data Distribution Application needs 12 language versions. For example, 12 xsl files are used for decoding and translation of XML export produced from the GEUS database.
- the name of each XSL contains an abbreviation of language
- the XSL file contains a full set of XML tags in accord with XSD (definition file)
- the XSL file serves to transform codes from WP5 XML into a selected language
- the structure of the transformed XML remains unchanged

7.2 PLSQL support for DMT export and MMT maintainance

As described under the heading “Data Model for Multilingual Thesaurus”, the structure of the MMT is relational. Branches, categories, codes, terms and their hierarchical relationships are stored separately. It is not possible to create a simple view of the database by constructing a spreadsheet necessary for the Distributed Multilingual Thesaurus (DMT). Therefore, several Oracle PLSQL packages were developed. Their purpose is: 1) to create a spreadsheet of

translated terms, and 2) to carry out integrity checks on the data and calculate statistics for the web application, which works with the Master Multilingual Thesaurus (MMT).

7.2.1 PLSQL package for DMT support

The DISTR_THES package performs a check on terms and their relationships, groups translations of terms according to keycodes, adds categories, descriptions of categories, and generates "ecode" (unique automatically generated eWater code), a detailed description of which is given under the heading "DMT export as XML table". This package is activated on demand, when XML export of Distributed Multilingual Thesaurus (DMT) is required.

7.2.2 PLSQL package for MMT support

This package is designed to support the MMT web application. The web application calls this package for communication with the Oracle database. This means that the web application does not use direct SQL commands but automatically calls procedures and functions for database maintenance. This kind of communication enables more complex querying/updating tasks and protects the MMT data from undesirable SQL command attacks. Moreover it separates middle tier database function from the user interface.

The main functions of the EWATER_STAT_W package are as follows:

- calculates statistics, creates output cursor tables for data presentations
- finds translated words
- creates cursor tables for all domain table presentations
- contains six pre-edit functions:
    - finds master terms still not translated into specified language
    - detects multiple translations of masters (one term to several master terms)
    - detects multiple translation to other languages (for example, two Dutch translations of one Czech term)
    - finds synonyms
    - finds autosynonyms
    - finds autotranslations
- creates support for viewing terms and their relationships
- performs some editing steps:
    - changes type of relationship (validate, delete or restore)
    - sets main relationship for term

8 Application for bulk loading of data and translation

8.1 Purpose of application

The first stage in the initiation of the MMT required large amounts of translated data to be loaded into Oracle tables. All these data are originally stored as MS Excel files organized according to national surveys and thematic branches. MS Excel format was chosen because of high readability and flexibility.

Because direct loading of data from MS Excel format into Oracle cannot be done, and because of the necessity for many logical operations during the loading process, a special tool for bulk load and translation was developed.

8.2 Basic functions

The newly developed application has three basic functions:

- a) bulk addition of new branches and categories
- b) bulk loading of new terms and master terms
- c) bulk translation of existing terms and comparison with master terms

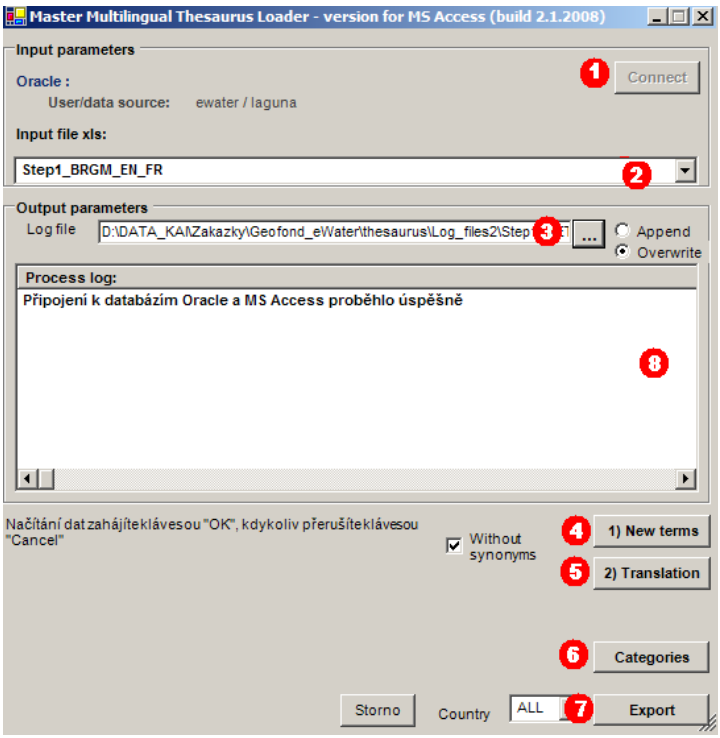
a) The MMT is designed as an open database, easily extended by addition of new branches and corresponding categories. As described in detail under the heading “Data model for the Multilingual Thesaurus”, the number of categories corresponds to the number of branches multiplied by the number of languages. Manual addition of new categories into the relational system involves much work. Therefore this process was automated and was incorporated in this application.

b) During the bulk loading process, new terms from the original MS Excel files and their master translations are loaded directly into the MMT tables. If terms already exist, their links (term ID’s) are used only to prevent data redundancy. Because the loading process is incremental, repetition of the same MS Excel files is possible. The process is repeated until no errors exist. All problems and errors are monitored in log-text files.

c) During the translation process, terms and their “master term” must exist (a strict requirement). The application loads translated terms from the original MS Excel file into the MMT and links these terms with the existing master and with terms that are to be translated. Simultaneously, all existing terms are checked against their “master term” and “automatic translations” (for different languages) or “automatic synonyms” (for the same language) are linked together. During this step, relationships of the type “explicit” or “automatic” translation are created, as described in detail under the heading “Master Multilingual Thesaurus”

Both loading and translation processes are transactional, therefore if any error arises, data are rolled back to preserve the relational integrity of the database. The Oracle RDBMS is responsible for these transactions.

The following Figure shows the "graphical user interface" of this application:



where user controls (enumerated with numbers in red circle) mean:

- (1) button for connection to the MMT
- (2) drop down list with input files (for example the file obtained from BRGM)
- (3) the browse button for selection of log file (file for monitoring the bulk loading process)
- (4) button for bulk loading of new terms and master terms
- (5) button for bulk translation of existing terms and comparison with master terms. Check box "Without synonyms" if checked, all recognized synonyms will be excluded (not stored)
- (6) button for automated generation of categories (if adding a new branch)
- (7) button for testing XML export to selected survey (country)
- (8) list box with status of the application

8.3 Development platform and technical parameters

The application version 2 was developed on the Microsoft Framework 2 platform (using the Microsoft Visual Studio .Net 2005 development tool) as a thick client of Windows. This platform fully supports Unicode encoding. Communication with the Oracle database is made directly using native oci.dll library, not by using the MDAC. The runtime version of this application needs a Windows platform with installed Microsoft Framework 2 software and Oracle Client software.

The Oracle 10g database for the MMT based on the Unix system platform is set to support UTF-8 encoding and is installed on the Internet database server in Geofond.

The application for bulk loading is installed on the client computer on a Windows platform with Microsoft Framework 2 software and with Oracle Client release 10.2.

9 Interactive internet application

9.1 Purpose of application

The web application is now designed to provide access to the MMT. This application permits the public user to search for hydrogeological or geological terms used in their national databases and to translate these terms into the working language of choice. This application also enables authorised users from the eWater consortium to review the status of the entries, to verify or, by using a simple edit function, to change the text of the translation. Statistical information on the entries in the thesaurus can be accessed both by the public and by authorised users, but at different levels of detail.
The newly developed application replaces the one designed for the existing geological thesaurus.

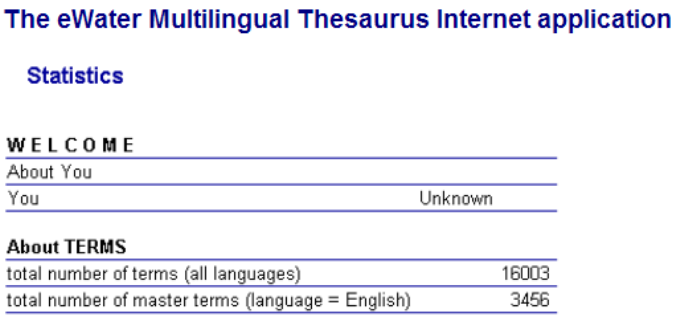
9.2 Basic functions

- Opening page
  - select the language for the GUI (graphical user interface)
  - link to eWater main portal and webpages of all participants

- Introductory page
  - a) starting an application for statistics (public access)
  - b) starting an application for a translation search (public access)
  - c) optional user authorization (login user) for advanced operations

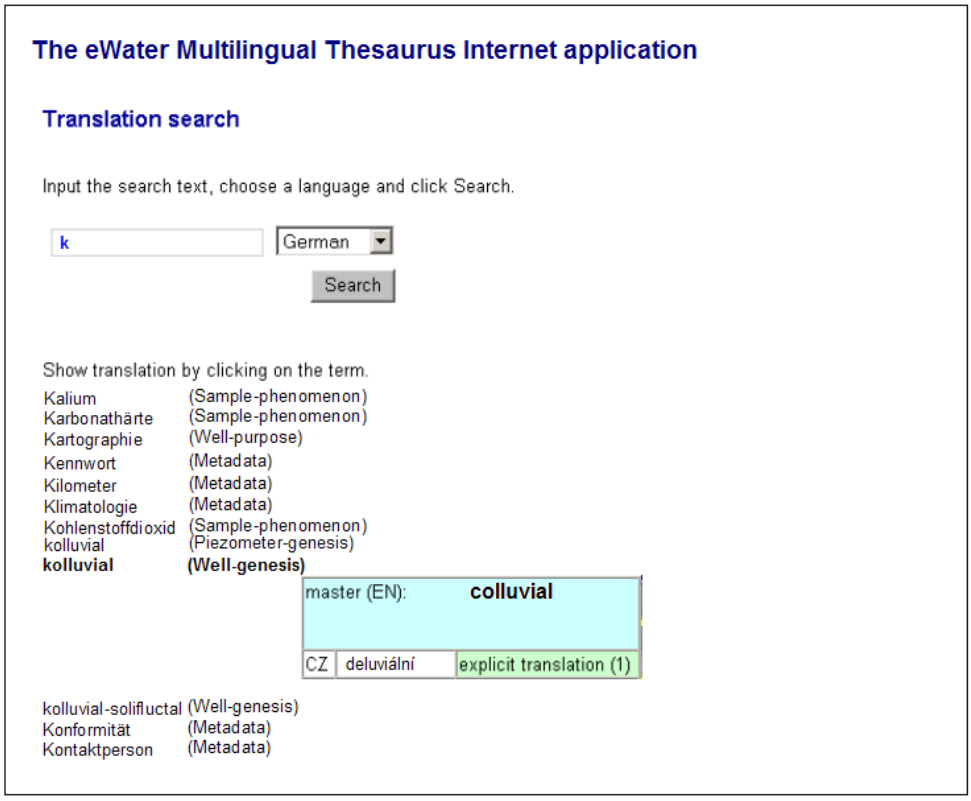
a) Statistics

The following figure shows the design of the page



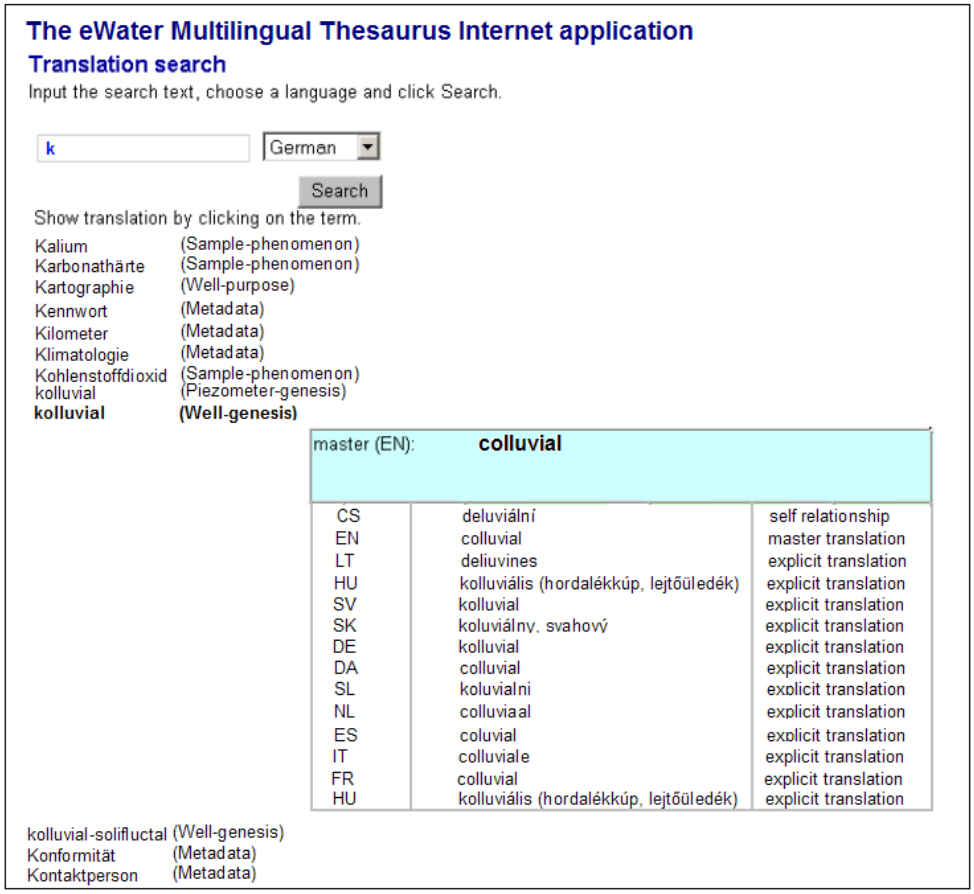
b) Translation search
- search for the list of terms using the implemented function for free-text search. If the text is prefixed by *, the search will yield terms in which search text is in any position.
- click on selected terms
- application generates table with translation

The following Figure shows user selection of all German terms, starting with "k". After clicking on "kolluvial", the application generates the table with translation into the Czech term "deluviální", also containing the master English term "colluvial" and the quality of translation - "explicit translation (1)"



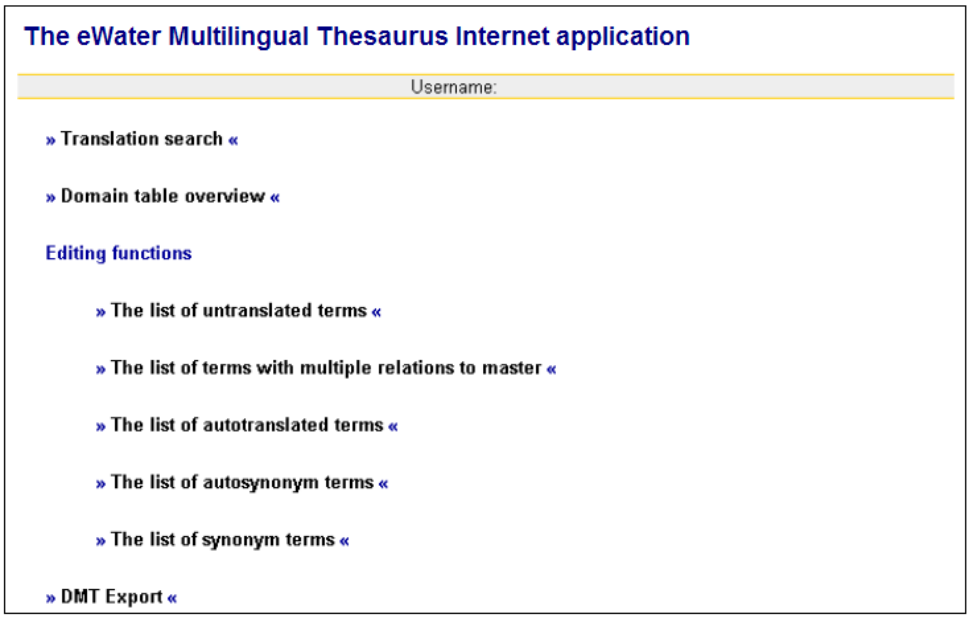
c) Authorized access. After successful user login, the "Translation search" dialog appears as shown in the following Figure. The dialog contains:

- indication of language selected
- an input form for the search text and selection of language
- list of terms found
- translation into all languages, including synonyms
- validated relationships (those validated by all members of the Consortium) visible to both authorized and public users
- non-validated relationships (those created automatically and for which validation by the members of the Consortium is pending) visible to authorized users only
- identity of the category to which the term belongs

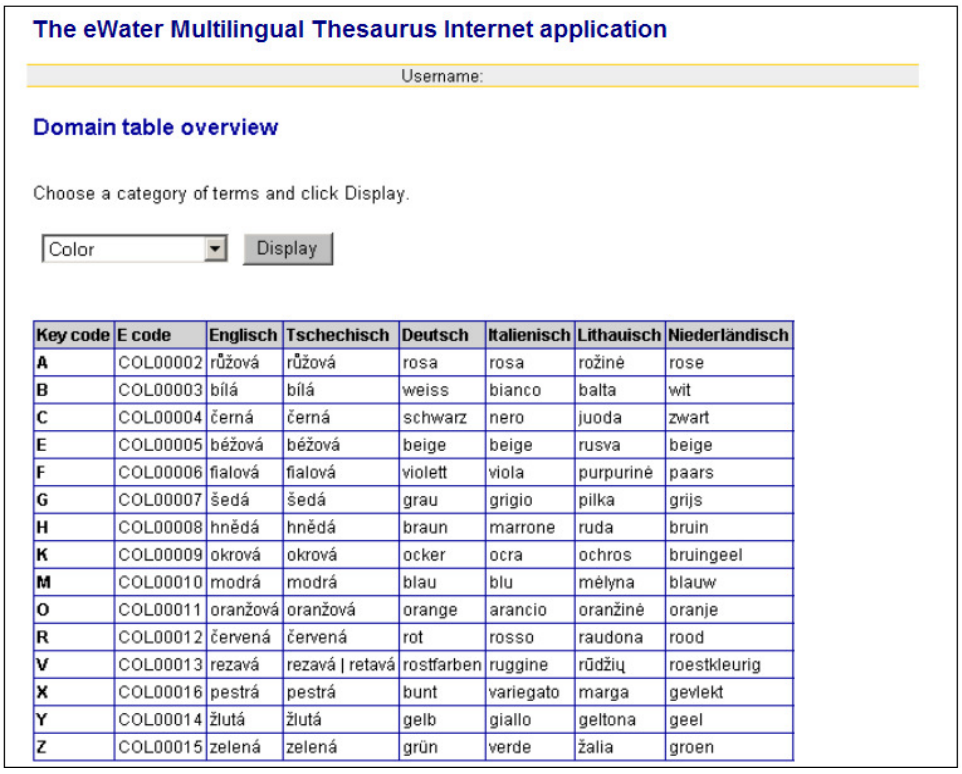


- Mode selection - authorized access. Function available to an authorized user:
  - reference to Statistics (extended)
  - reference to Domain table overview
  - reference to Editing functions
  - reference to DMT Export

The following Figure shows a menu with a choice of possible operations:



- Overview of the domain table (table containing terms from one branch) for authorized access
  - an input form for selection of category
  - domain table overview in all languages



The eWater Multilingual Thesaurus Internet application

Username:

Domain table overview

Choose a category of terms and click Display.

Color | Display

Key code	E code	Englisch	Tschechisch	Deutsch	Italienisch	Lithauisch	Niederländisch
A	COL00002	růžová	růžová	rosa	rosa	rožinė	rose
B	COL00003	bílá	bílá	weiss	bianco	balta	wit
C	COL00004	černá	černá	schwarz	nero	juoda	zwart
E	COL00005	béžová	béžová	beige	beige	rusva	beige
F	COL00006	fialová	fialová	violett	viola	purpurinė	paars
G	COL00007	šedá	šedá	grau	grigio	pilka	grijs
H	COL00008	hnědá	hnědá	braun	marrone	ruda	bruin
K	COL00009	okrová	okrová	ocker	ocra	ochros	bruingeel
M	COL00010	modrá	modrá	blau	blu	mėlyna	blauw
O	COL00011	oranžová	oranžová	orange	arancio	oranžinė	oranje
R	COL00012	červená	červená	rot	rosso	raudona	rood
V	COL00013	rezavá	rezavá | retavá	rostfarben	ruggine	rūdžių	roestkleurig
X	COL00016	pestrá	pestrá	bunt	variegato	marga	gevlekt
Y	COL00014	žlutá	žlutá	gelb	giallo	geltona	geel
Z	COL00015	zelená	zelená	grün	verde	žalia	groen

9.3 Development platform and technical parameters

The internet application for MMT review and simple editing of data is now being developed for Microsoft Framework 2 platform (using the Microsoft Visual Studio .Net 2005 as a development tool). ASP.NET application communicates with Oracle database via Oracle Client v.10.2. The Microsoft Internet Information Server is used as the internet server.

10 Results and Conclusions

Principal aims of the WP 8
- a) to establish and fill the MMT with terms, metadata, headings and user interface terms
- b) to provide exports from the MMT in the form of a DMT: for the DDA applications and the Central Portal application
- c) to allow distant MMT maintenance
- d) to provide a translational multilingual thesaurus in hydrogeology and geology to the internet public

The state of completion of the individual tasks:
ad a)
- the MMT database was established and filled with most terms used in national databases; the terms were translated to the national languages
- the MMT was filled with metadata terms, translated into all languages
- the MMT was filled with explanatory texts, translated into all languages
- the translation of four user interface files is in progress, user interface terms have not been loaded into the MMT yet - As the process of creation of applications is still ongoing, this set of terms will probably be modified or extended depending on the parts of the applications ratified during the phase of implementation.

ad b)
- the export of the DMT in the XML format was managed and realized
- the export of the DMT in the XSL format is in progress; it is designed to be compatible with the DDA application developed by the GEUS within WP7 and 9

ad c), ad d)
- a new version of the application for internet access to the MMT is being developed. This application will provide public interactive access to the hydrogeological and geological dictionaries. For an authorized user, the application will provide information on the reliability of the translation and on the state of filling the MMT; it will also allow simple editing of the translations of terms. The final version of the application will be probably modified depending on the ratified portal applications during the Implementation phase.

The developed dictionary has been serving for eWater system development and will insure the system multilinguality upon its completion.

11 Bibliography and References

http://www.cig.ensmp.fr/~hydro/BIB/UNESCO/TDH000.pdf, Intl. Glossary of Hydrogeology, UNESCO 1978

Domenico, P.A., Schwartz, F.W. Physical and Chemical Hydrogeology, 2nd Ed., John Wiley and Sons, New York. 1998

Jackson, J.A. (1997): Glossary of Geology. 4th Ed., Amer. Geol. Inst., Alexandria, Virginia, USA. 769 pp.

http://www.groundwater.org/gi/gwglossary.html, The Groundwater Foundation

http://water.nv.gov/water%20planning/dict-1/ww-index.cfm, Nevada Division of Water Planning (Gary A. Horton)

http://www.if.uidaho.edu/~johnson/ifiwrri/sr3/gloss.html

http://www.brown.edu/Courses/GE0158/web2_revised/webglossary/hdef/hydrogeology.html

Annexes

A1. XML Schema Definition file XSD for Distributed Multilingual Thesaurus

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema
	xmlns="http://www.ewater.eu/mmt2dmt"
	xmlns:xs="http://www.w3.org/2001/XMLSchema"
	targetNamespace="http://www.ewater.eu/mmt2dmt"
	version="1.0">
	<xs:annotation>
		<xs:documentation><![CDATA[
			Schema for XML export from Master Multilingual Thesaurus (MMT)
			GEOFOND (2007)

			XML export file contains only terms for ONE national database.
			Terms are translated into all languages. Quality of translation is evaluated with attribute "qual".
			Each term is translated as a couple:
				UTF-8 text
				HTML representaion of text with special characters (&#nnn;)
			Terms are organized into categories (branches)
		]]></xs:documentation>
	</xs:annotation>
	<xs:element name="terms">
		<xs:annotation>
			<xs:documentation>
				repeating tag for each term
			</xs:documentation>
		</xs:annotation>
		<xs:complexType>
			<xs:attribute name="cat" use="required">
				<xs:annotation>
					<xs:documentation>
						translated name of category
					</xs:documentation>
				</xs:annotation>
				<xs:simpleType>
					<xs:restriction base="xs:string"/>
				</xs:simpleType>
			</xs:attribute>
			<xs:attribute name="catid" use="required">
				<xs:annotation>
					<xs:documentation>
						acronyme (code) for category (branch)
					</xs:documentation>
				</xs:annotation>
				<xs:simpleType>
					<xs:restriction base="xs:integer"/>
				</xs:simpleType>
			</xs:attribute>
			<xs:attribute name="ecode" use="required">
				<xs:annotation>
					<xs:documentation>
						unique code for each national survey term (generated inside MMT)
					</xs:documentation>
				</xs:annotation>
				<xs:simpleType>
					<xs:restriction base="xs:string">
						<xs:pattern value="[A-Z]+[0-9]+"/>
					</xs:restriction>
				</xs:simpleType>
			</xs:attribute>
			<xs:attribute name="hterm" use="required">
				<xs:annotation>
					<xs:documentation><![CDATA[
```


```
                                        translated term with special characters (d&#nnn;) as HTML expressions
                                ]]></xs:documentation>
                        </xs:annotation>
                        <xs:simpleType>
                                <xs:restriction base="xs:string"/>
                        </xs:simpleType>
                </xs:attribute>
                <xs:attribute name="lang" use="required">
                        <xs:annotation>
                                <xs:documentation>
                                        code for all required languages
                                </xs:documentation>
                        </xs:annotation>
                        <xs:simpleType>
                                <xs:restriction base="xs:string">
                                        <xs:enumeration value="CS"/>
                                        <xs:enumeration value="DA"/>
                                        <xs:enumeration value="DE"/>
                                        <xs:enumeration value="EN"/>
                                        <xs:enumeration value="ES"/>
                                        <xs:enumeration value="FR"/>
                                        <xs:enumeration value="HU"/>
                                        <xs:enumeration value="IT"/>
                                        <xs:enumeration value="LT"/>
                                        <xs:enumeration value="NL"/>
                                        <xs:enumeration value="SK"/>
                                        <xs:enumeration value="SL"/>
                                        <xs:enumeration value="SV"/>
                                </xs:restriction>
                        </xs:simpleType>
                </xs:attribute>
                <xs:attribute name="ncode" use="required">
                        <xs:annotation>
                                <xs:documentation>
                                        national survey code (optionally for terms with coded values)
                                </xs:documentation>
                        </xs:annotation>
                        <xs:simpleType>
                                <xs:restriction base="xs:string"/>
                        </xs:simpleType>
                </xs:attribute>
                <xs:attribute name="qual" use="required">
                        <xs:annotation>
                                <xs:documentation>
                                        code for quality of translation:
                                                1 = explicit (user confirmed translation)
                                                2 = master translation (english translation)
                                                3 = auto translation (automated translation made by MMT)
                                                9 = self relationship (national survery term, which is
translated)
                                                11 = explicit synonym (user confirmed synonym)
                                                13 = auto synonym (synonym created automatically)
                                </xs:documentation>
                        </xs:annotation>
                        <xs:simpleType>
                                <xs:restriction base="xs:integer">
                                        <xs:enumeration value="1"/>
                                        <xs:enumeration value="2"/>
                                        <xs:enumeration value="3"/>
                                        <xs:enumeration value="9"/>
                                        <xs:enumeration value="11"/>
                                        <xs:enumeration value="13"/>
                                </xs:restriction>
                        </xs:simpleType>
                </xs:attribute>
                <xs:attribute name="source" use="required">
                        <xs:annotation>
                                <xs:documentation>
                                        optional code of national survey data source (e.g. bh=borehole database)
                                </xs:documentation>
                        </xs:annotation>
                        <xs:simpleType>
                                <xs:restriction base="xs:string"/>
                        </xs:simpleType>
```

```
                        </xs:attribute>
                        <xs:attribute name="term" use="required">
                                <xs:annotation>
                                        <xs:documentation>
                                                translated term with special characters as UTF-8
                                        </xs:documentation>
                                </xs:annotation>
                                <xs:simpleType>
                                        <xs:restriction base="xs:string"/>
                                </xs:simpleType>
                        </xs:attribute>
                </xs:complexType>
        </xs:element>
        <xs:element name="translation">
                <xs:annotation>
                        <xs:documentation>
                                root element containing terms
                        </xs:documentation>
                </xs:annotation>
                <xs:complexType>
                        <xs:sequence>
                                <xs:element ref="terms" maxOccurs="unbounded"/>
                        </xs:sequence>
                        <xs:attribute name="survey" use="required">
                                <xs:annotation>
                                        <xs:documentation>
                                                national survey codes
                                        </xs:documentation>
                                </xs:annotation>
                                <xs:simpleType>
                                        <xs:restriction base="xs:string">
                                                <xs:enumeration value="BRGM"/>
                                                <xs:enumeration value="GEOFOND"/>
                                                <xs:enumeration value="GEOZS"/>
                                                <xs:enumeration value="GEUS"/>
                                                <xs:enumeration value="GPA"/>
                                                <xs:enumeration value="GSSR"/>
                                                <xs:enumeration value="IGME"/>
                                                <xs:enumeration value="LGT"/>
                                                <xs:enumeration value="MAFI"/>
                                                <xs:enumeration value="SGSS"/>
                                                <xs:enumeration value="SGU"/>
                                                <xs:enumeration value="TNO"/>
                                        </xs:restriction>
                                </xs:simpleType>
                        </xs:attribute>
                </xs:complexType>
        </xs:element>
</xs:schema>
```