

# **Assembly of Gemet 2000**

**Report on the assembly of Gemet 2000 including the integration into THESshow**

**August 2000**

**Hermann Stallbaumer**

## Preparing the database

The creation of the Gemet 2000 database started from the Gemet 2.0 database from August 1999.

The following steps were performed to enable the import of the additional language data:

- Rearrangement of languages
- Moving of terms
- Adaption of Nondescnumbering
- Recreation of hierarchy tables

### Language arrangement in Gemet 2000

Lang	Language	Font	Ord	Abbrev	Codepage	Charset
1	English	Times New Roman	1	eng	0	Iso-Latin 1
2	Danish	Times New Roman	2	dan	0	Iso-Latin 1
3	Finnish	Times New Roman	3	fin	0	Iso-Latin 1
4	German	Times New Roman	4	ger	0	Iso-Latin 1
5	Dutch	Times New Roman	5	nld	0	Iso-Latin 1
6	Norwegian	Times New Roman	6	nor	0	Iso-Latin 1
7	Swedish	Times New Roman	7	swe	0	Iso-Latin 1
8	French	Times New Roman	8	fre	0	Iso-Latin 1
9	Greek	Times New Roman	9	grk	161	Iso-Latin 7
10	Italian	Times New Roman	10	ita	0	Iso-Latin 1
11	Portuguese	Times New Roman	11	por	0	Iso-Latin 1
12	Spanish	Times New Roman	12	spa	0	Iso-Latin 1
13	Hungarian	Times New Roman	13	hun	238	Iso-Latin 2
14	Slovak	Times New Roman	14	slk	238	Iso-Latin 2
15	US-English	Times New Roman	15	usa	0	Iso-Latin 1

## Import of Languages

### Hungarian:

Groups:

40 translated groups were delivered, 39 imported (*Not accepted terms*, which is not part of Gemet was not imported)

Themes:

40 were delivered, 40 were imported.

Terms:

5311 terms delivered, 5295 imported. Not imported were the 16 terms from Group NAT which is omitted in Gemet-2000.

Terms of Group NAT
Hearing authority
Cantonal law
federal highway
plant size (industry)
DIN standard
mineral rock and earth industry
feed-in current
Brackish
space utilisation
Thickness of insulating layer
MSR technique
textile technology
feeding current
right of access to administrative
noise impact in daily life
statement of facts

1 term was left untranslated.

DescNr	Desc1
2381	Dust

In the delivered file 4328 terms have an identifier of 2291. So the data synchronisation was only possible by english terms. No back check was possible.

## Slovak:

### Groups:

40 translated groups were delivered, 39 imported (*Not accepted terms*, which is not part of Gemet was not imported)

### Themes:

40 were delivered, 40 were imported.

### Terms:

5311 terms were delivered, 5295 were imported into Gemet 2000. Again those from the Group *NAT* were not imported.

69 terms (67) are still missing

DescNr	Desc English
986	Bradyseism
10071	sand flat
9019	Wadden Sea
10146	Geographic circque
10043	plant life
10003	Animal life
755	Beaching
13639	Bovine
8286	tank farm
12170	Ocean outfall
428	Animal housing
9236	Waterside development
10243	Classified site
5937	Organotin compound
87	Addition polymer
5701	novel food
6828	Pulp
8140	Strength of materials
13193	Agricultural holding
1725	Construction of installations
7897	soil surface sealing
12525	Timber producing chain
13219	Materials technology
8350	Telematics
1763	Contract cleaner
12379	Health-care activities waste
9132	Wastewater from trade
9045	Waste assimilation capacity
6473	Pollution sink
5571	net resource depletion
1291	Negotiable charge
1822	cost recovery basis
11073	antipollution premium
1446	class action suits law
14966	natural independence law
4715	neighbourhood law
2226	discharge legislation
4798	licencing obligation

2769	environmental protection advice
654	attribution
1663	compulsory use
11007	motivation of administrative acts
1229	cause for concern principle
1280	chain management
11568	administrative court (administration)
6293	Planning-Programming-Budgeting System
12459	classified facility
11248	equipment plan
8562	tracking plan
8145	structural adjustment program
4921	lower risk species (IUCN)
2887	environmental performance
10251	urban ecology charter
5093	materials science
11359	municipal engineering
6791	public emergency limit
8411	testing guideline
5863	olfactometry
11812	drawing
2219	disaster control service
3611	gender issue
1928	custom and usage
5783	occupational group
9409	youth work
1022	Brussels additional agreement
4796	licencing
5428	municipal cleansing

## Danish

### Groups:

39 translated groups were delivered, 39 imported

### Themes:

40 were delivered, 40 were imported.

### Terms:

4871 descriptors were imported from the database Master-21.

In the new file 409 danish descriptors were delivered, 406 were imported.

Three terms could not be imported:

electric power network	13592	(Could not be located)
environmental economy	14298	(Nondescriptor)
additive effect of toxic substances	89	(Nondescriptor)

18 descriptors are still missing.

Many (947 ) entries are multiple terms divided by „;“ They were disassembled. The first token was used as a descriptor, the following entry or entries were classed as nondescriptors.  
1283 nondescriptors were created.

## Import of Definitions

1121 delivered by US EPA,

1115 imported,

Not imported are

ENG	ID	Reason
immission control legislation	14444	Nondescriptor
Agreement	13728	Nondescriptor
plant size (industry)	6322	Not in Gemet 2.0
right of access to administrative documents	11529	Nondescriptor
Remuneration of work	13243	Nondescriptor
Reuse	7200	Nondescriptor

The abbreviations in Ref\_ENG were converted in (Source: <REF\_ENG>) and appended to the definitions. Commas, plus-signs and ampersands were replaced by Foreslashes.

Definitions already present were overwritten.

All entries in Definitions „No definition needed, (Source: NDN)“ were replaced by „No definition needed“

Unfortunately the EPA did not deliver a description for the abbreviations. A part of it is already used in Gemet 2.0

#### List of used abbreviations.

AAG	EPW	LES	RIC
ADM	ERD	LFS	RMT
AFR	ERG	LIV	ROA
AGP	EUR	LLL	RSG
ALL	EWP	LTM	RUR
ANT	FAI	LWW	SAR
APD	FEM	MED	SCP
APS	FFD	MES	SHG
AUB	FIT	MGHME	SIC
AUS	FOI	MHD	SIS
AVA	FPL	MHE	SMD
AWD	FRE	MIB	SOC
BAS	FUT	MLK	SOI
BHE	GAB	MLM	SPR
BIOSAF	GAS	MSE	SRD
BLD	GET	MTE	TED
BWP	GHT	MUM	TEF
C	GIS	MVG	TEX
CBO	GLG	NAH	TIM
CCL	GMR	NCF	TOE
CED	Greenwald	NDECON	tr Z
CEN	GRP	NGT	TWI
CGP	GRT	NIO	TWS
CIA	GSW	NOA	UIA
CIV	GT2	NOV	UMT
CNI	GTT	NPH	Unep
COE	GWG	NRW	Unesco
CON	HEG	NTL	UNM
CSD	HGD	OCT	USC
CTC	HHR	ODE	VFP
DAM	HMD	OEC	WAE
DDP	HMH	OED	WAP
DEE	HMO	OER	WBG
DEP	HRE	OHS	WCD
DES	HUD	OMD	West's
DHW	HUM	OPP	West'sX
DMG	IEC	ORB	WHIT
DOE	IEL	OSW	WIC
DOG	INP	OXF	WIK
DPG	INW	PAJ	WIS
DWT	IPC	PBS	WMA
ECH	IPE	PEM	WON
ECW	ISEP	PGE	WOR
EEN	IVW	PLA	WQA
EFP	JON	PMA	WQL
EIS	JSS	POP	WRT
ELD	KIN	PPP	WWC
ENA	KSA	PVG	WWD
ENC	LAB	QAG	
ENZ	Landy	RAD	
EOC	LCC	RHW	

## Changes in the French data

Due to customer requests the following french entries were changed:

"évaluation de l'impact sur l'environnement" -> "évaluation..."

"émissions industrielles" -> "émissions..."

"composé organohalogène" -> "composé organohalogéné"

Not possible, for it would create duplicate entries.

The term

"rétonnée atmosphérique" should write "retombée atmosphérique"

This is not possible, for another term "retombée atmosphérique" is already present. A change would create a duplicate entry, which is not possible. To overcome this problem, either a different term or a qualifier in parantheses could be added.

Also changed:

„évaluation des risques pour l'environnement“

The next errors could not be found in the Gemet 2.0 THESmain database nor in the Gemet 2.0 THESshow database! The error is most likely a translation error from or to SGML

construction &#034;verte&#34;  
architecture &#034;durable&#034;  
construction &#034;verte&#034;  
procédé &#034;Thermosélect&#034;  
instruments de fiscalité &#034;verte&#034;  
politique de la &#034;bulle&#034; (permis d'émissions négociables)  
technologie &#034;end-of-pipe&#034;  
technologie &#034;propre&#034;

## THESshow Icon for Winword

If Winword is registered as default text system on a computer, the installation program of THESshow allows to create an icon in Winword in order to launch THESshow. If both applications are running, thesaurus daty may be copied from THESshow to Winword by Drag and Drop. Of course legacy Cut and Paste is possible also.

The feature works for Winword 97 and Winword 2000 (both are tested). When deinstalling THESshow, the icon is also removed.