



<http://intevation.net>

**GREAT-ER II**

# **GREAT-ER Preprocessing Manual**

Intevation GmbH  
Georgstraße 4  
D-49074 Osnabrück  
Germany

Date: 17.Jul.2003

This document has been designed with  $\text{\LaTeX}$ . It is available as source code, PDF- and HTML-format.

Version: 1.2.0

Date: 17.Jul.2003

Authors: Jan-Oliver Wagner · Frank Koormann

Copyright (c) 2002, 2003 Intevation GmbH.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	History . . . . .	5
1.2	Requirements . . . . .	6
<b>2</b>	<b>How-To</b>	<b>7</b>
2.1	Software Requirements . . . . .	7
2.2	Prerequisites . . . . .	7
2.3	Running the Data Preprocessing . . . . .	8
2.4	Integrate a catchment into GREAT-ER 1.0 . . . . .	8
2.5	Integrate a catchment into GREAT-ER 2.0 . . . . .	8
<b>3</b>	<b>Catchment Source Files Formats</b>	<b>9</b>
3.1	Info File, File 'catchment.desc' . . . . .	9
3.2	Digital River Network, File 'ID.drn' . . . . .	9
3.3	River Network Attributes, File 'ID.rna' . . . . .	10
3.4	Discharge Site Data, File 'ID.dsd' . . . . .	10
3.5	Catchment Boundary Polygon, File 'ID.cbp' . . . . .	11
3.6	Lakes, File 'ID.lks' . . . . .	11
3.7	Pictures, File 'ID.pic' . . . . .	11
3.8	Background Data, File '.bgd' . . . . .	12
<b>4</b>	<b>Catchment Result Files Formats</b>	<b>15</b>
4.1	Map Layers . . . . .	15
4.2	River Network Attributes, File river.dbf . . . . .	16
4.3	Discharge Site Data, File disch.dbf . . . . .	16
4.4	Discharge Classes, File wwtp.dbf . . . . .	17
4.5	River Classes, File rivclass.dbf . . . . .	17
4.6	Background Data List, File backgroundlst.dbf . . . . .	18
<b>5</b>	<b>Preprocessing Basics</b>	<b>19</b>
5.1	Data requirements and composition . . . . .	22
5.1.1	Digital river network . . . . .	22
5.1.2	River network attributes . . . . .	24
5.1.3	Catchment boundary . . . . .	28
5.1.4	Discharge data . . . . .	28

5.1.5	Background maps . . . . .	30
5.2	Final automatic data processing . . . . .	31
5.2.1	Geographic data . . . . .	31
5.2.2	Attribute data . . . . .	32
5.2.3	Core algorithms . . . . .	32
	<b>References</b>	<b>37</b>
	<b>A GNU Free Documentation License</b>	<b>39</b>

# 1 Introduction

The GREAT-ER pre-processing is a script-based routine to prepare a GREAT-ER catchment dataset from raw data in a pre-defined text format.

This document intends to enable the reader to prepare catchments for use with the GREAT-ER simulation system. The process as a whole is described as well as technical details on how the procedure works and on how to apply the associated software tools that actually perform the processing. The commented source code itself (released as Free Software) completes the documentation on all details.

Current information, all source codes, documentation and a mailing list can be found at the GREAT-ER development homepage:

<http://great-er.intevation.org/>

## 1.1 History

The pre-processing routines originate from the GREAT-ER 1 project as a script collection and have been further improved since then.

The major changes from GREAT-ER version 1.0 to version 2.0 are:

- Availability as Free Software.
- Independency from proprietary software (especially ArcInfo which was a major drawback due to its high license costs).
- Replacement of ArcInfo coverage format by ESRI Shapefile format for all geographic data.

The detailed changes of the present version are listed in the file `Changes.txt` of the pre-processing software package.

## 1.2 Requirements

Applying the GREAT-ER pre-processing requires experience with Unix-like systems; principle programming skills are helpful.

The basic requirements are

- a usual Unix-like system including the standard GNU software tools (recommended is a GNU/Linux system).
- The programming language Python which is usually available on such systems, but they may not be updated to Python version 2.0 as required for the pre-processing.
- The special tools `gen2shp`, `txt2dbf` and `dbf2txt`.

Details, e.g. on specific versions on the required tools are given in the Howto-section (see below).

## 2 How-To

### 2.1 Software Requirements

- Python  $\geq$  2.2.1
- GNU Awk  $\geq$  3.1.0
- gen2shp  $\geq$  0.3.0  
(<http://intevation.de/~jan/gen2shp/>)
- txt2dbf  $\geq$  1.0.2
- dbf2txt  $\geq$  0.5.2  
(<http://www.usf.uos.de/~fkoorman/software/dbftools.en.html>)
- GNU textutils (sort, join, cut, paste)  $\geq$  2.0
- GNU fileutils (rm)  $\geq$  4.1
- GNU sed version  $\geq$  3.02
- GNU sh-utils (echo)  $\geq$  2.0.11

(all the programs must be somewhere in your search path)

### 2.2 Prerequisites

As a first step you have to copy greater-pre-processing.py somewhere into you PATH or extend the PATH to where you have installed the greater-pre-processing package:

```
cp greater-pre-processing.py /some/where/in/your/search/path
```

OR

```
export PATH=$PATH:/some/where/greater-pre-processing
```

Next, set this environment variable that points to where you have installed the greater-pre-processing package:

```
export GREATER_PRE_PROC=/some/where/greater-pre-processing
```

You then are able to execute

```
greater-pre-processing.py
```

in a directory with the pre-defined files for your catchment (you will be given further info on how to use the program).

## 2.3 Running the Data Preprocessing

```
$ greater-pre-processing.py help
greater-pre-processing 1.2.0
```

```
Found 'catchment.desc':
  ID                = aire
  Name              = Catchment of River Aire including River Calder
  Description       = The Aire catchment is in Yorkshire, UK
  Format version src = 1.0
  Format version result = 2.0
description complete.
```

```
greater-pre-processing [check|help|run|run-without-cleanup|clean]
  help print this text
  check test for completeness and correct syntax of source files
  run execute catchment pre-processing
  run-without-cleanup execute catchment pre-processing not removing
  temporary files
  clean remove all temporary and all results files
```

## 2.4 Integrate a catchment into GREAT-ER 1.0

First, make sure, that OUT-VERSION in catchment.desc is set to 1.0, then process the data.

Now copy all files \*.shp, \*.shx, \*.dbf into %GHOME%/catchments/ID where ID is the name of the new catchment. Finally, add the new catchment to %GHOME%/catchments.dbf, see [http://great-er.intevation.org/catchments\\_dbf.html](http://great-er.intevation.org/catchments_dbf.html) for how to do this.

## 2.5 Integrate a catchment into GREAT-ER 2.0

First, make sure, that OUT-VERSION in catchment.desc is set to 2.0, then process the data.

With the GREAT-ER II Administration Tool:

- Create a new catchment entry (CATCH\_TAB)
- Upload all geographic data into the data base (BIN\_OBJ\_TAB)
- Relate available background data in the database to the new catchment (BACKGRD\_JOIN\_TAB)

Prepare the discharge and stretch attributes by running the scripts `insert-disch_att.py` and `insert-stretch_att.py`. The scripts generate sequences of SQL-statements. These can be executed by the database SQL shell to upload the attributes.

## 3 Catchment Source Files Formats

*Note:* This new format slightly differs from the original GREAT-ER 1 format and is therefore not fully compatible. See the file *Changes.txt* of the software package for details.

All 6 catchment source files have to be collected in one directory.

All files are text based and use the character set ISO 8859-1 (ISO Latin 1). The format is line-oriented; each line starts with a '#' (comment line) and all empty lines have no meaning for the processing and can be inserted anywhere.

All coordinates are given in the same projection: "geographic projection" with unit "dd" (decimal degrees). A point, rather than a comma, is used as decimal point. Numbers are not allowed to contain any other delimiters. Exponential representation uses the letter 'E' or 'e'.

### 3.1 Info File, File 'catchment.desc'

This file contains general information. Example:

```
ID=went
NAME=Catchment of River Went
DESCRIPTION=The Went catchment is a small, rural area in Yorkshire, UK.
IN-VERSION=1.0
OUT-VERSION=1.0
```

**ID** is the basename for the actual data source files suffixed with .drn, .rna, .dsd, .cbp and .bgd.

**IN-VERSION** defines the format for the all catchment source files.

**OUT-VERSION** defines the output format.

### 3.2 Digital River Network, File 'ID.drn'

IN-VERSION = 1.0

The river network attributes are stored in a simple tabular structure. Columns are separated by commas. The following example illustrates the order of the columns:

```
7421
-1.399004,53.718807
-1.374895,53.730835
```

```
-1.368942,53.734528
-1.364163,53.736794
-1.363446,53.738102
6787
-1.399046,53.717968
-1.399043,53.718178
[...]
```

### 3.3 River Network Attributes, File 'ID.rna'

IN-VERSION = 1.0

The river network attributes are stored in a simple tabular structure. Columns are separated by commas. The following example illustrates the order of the columns:

```
# Aire/Calder Catchment, Yorkshire
# StretchID,Qmean(m^3/s),Q5(m^3/s),vmean(m/s),v5(m/s),
# ... RealLength(m),depthmean(m),depth5(m),Name
26954,34.751000,12.280656,,,2491.655000,,,Unnamed
22954,0.009000,0.002359,,,1648.222000,,,Unnamed
7333,34.739000,12.276415,,,1629.885000,,,Unnamed
22932,0.005000,0.001088,,,540.190000,,,Unnamed
26950,34.731000,12.274977,,,2336.740000,,,Unnamed
22923,0.017000,0.002167,,,2447.520000,,,Unnamed
[...]
```

In this example no information is available for flow velocity and depth and hence the corresponding columns are left empty.

Note: In case of a lake, the field for RealLength should contain the volume in  $m^3$ !

### 3.4 Discharge Site Data, File 'ID.dsd'

IN-VERSION = 1.0

Each line defines one discharge. The first column is its unique ID, followed by the coordinate pair and the attributes. The order of the columns is shown in this example:

```
# ID,X,Y,Pop,DWF(m^3/d),Flow(m^3/d),Type,StretchID,Name
41,-1.390494,53.422081,422677,130000,180000,AS,26842,Blackburn STW
42,-1.349155,53.369724,95750,21800,24506,AS,26843,Woodhouse Mill STW
43,-1.326922,53.445114,107203,31250,32410,AS,26841,Aldwarke STW
```

### 3.5 Catchment Boundary Polygon, File 'ID.cbp'

IN-VERSION = 1.0

The polygon is described linewise by single coordinate pairs (x,y). The following example illustrates the format:

```
-2.052912,53.683811  
-2.060483,53.683807  
-2.060477,53.679314  
-2.068047,53.679310  
[...]  
-2.045346,53.688309  
-2.045342,53.683804  
-2.052912,53.683811
```

This example also shows that the polygon must be closed, cf. the last and the first point must be identical.

### 3.6 Lakes, File 'ID.lks'

IN-VERSION = 1.0

A list of stretch IDs which actually are lakes. Example:

```
# StretchID  
345  
2256
```

This file can be empty, but must exist.

### 3.7 Pictures, File 'ID.pic'

IN-VERSION = 1.0

If pictures are available for the catchment, location and picture files can be specified:

```
# ID,X,Y,File,Name  
1,6.860403,51.157661,mgmuendung.jpg,Confluence to the Rhine  
3,6.872726,51.158089,mgddorf.jpg,in Duesseldorf  
4,6.897292,51.159611,uhilden1.jpg,Below STP Hilden  
5,6.899222,51.159081,uhilden2.jpg,Below STP Hilden
```

This file can be empty, but must exist.

**Note**, that the picture files must be in jpg-format.

## 3.8 Background Data, File '.bgd'

IN-VERSION = 1.0 (however, you can leave this file empty for GREAT-ER 2.0)

To support the user with a quick and easy access to geographic background information, a list of background data sets can be defined for each catchment.

In the ".bgd" file names of additional background data that are to be loaded by the menu entry "Display/Add Background Data" are specified. The data can be given in two ways. In a simple case, a background data set is only a geographic data set that ArcView can load as a theme, i.e. an ArcInfo Coverage, an image, a grid or a shapefile. This case is indicated by the value "yes" in the field "ctch\_flg" (see below). In the more complex case a background data set consists of a set of geographic data, indicated by the value 'no' in the field 'ctch\_flg'. In this case, instead of the name of the data set an ID for a script that knows how to load the geo-data sets is specified.

### Name

This is the name for the corresponding background data. It is used for user selections in the GREAT-ER system.

### Mapname

In the case of a simple background data set, this is the file name of the coverage, image or shapefile that contains the data. Otherwise this field contains the script which can load the background data sets.

### Type

In the case of a simple background data set, this is its feature type. Valid entries are "point", "line", "polygon", "shape", "image" and "grid". Otherwise the field is left empty.

### Legendfile

In the case of a simple background data set, this is the name of an optional ArcView legend file. Otherwise this field can contain (optional) parameters for the script specified in the field "Mapname".

### ctch\_flg

This is a boolean field that contains either "yes" or "no". "yes" indicates that this background data set consists of one theme that does not need a special script to load into the active view of GREAT-ER. "no" means that a special Avenue script has been developed by the user which runs within GREAT-ER and loads the background data set into the active view.

There is one special case: The DCW data (Digital Chart of the World, a variant of VMAPL0) are incorporated in the GREAT-ER project for the whole of Europe. For this incorporation specific

scripts loading the DCW data in dependence of the viewed country are developed. If the DCW data are to be a background data set for a catchment, the following line needs to be inserted into the file:

```
DCW data,,,,
```

If a user installs the DCW data for countries that have not yet been included, e.g. France, Spain or Denmark, these data need to be installed analogous to the included countries.

### **Example ".bgd" file**

This example file describes background information for the River Itter (Germany):

```
#River Itter background data
#Name,Mapname,Type,Legendfile,ctch_fl
Cities,itter\cities,polygon,itter\cities.avl,yes
Tk25,itter\tk25tiffz.tif,image,,yes
Rhein,itter\rhein,polygon,itter\rhein.avl,yes
DCW data,,,,no
```



## 4 Catchment Result Files Formats

Currently two versions of the results files can be created:

OUT-VERSION = 1.0: creates catchment datasets compatible with GREAT-ER 1.

OUT-VERSION = 2.0: creates catchment datasets compatible with GREAT-ER 2. The difference is additional columns `down1` and `down2` in `river.dbf` and different Stretch-Types for lakes. Furthermore, no background file `backgroundlst.dbf` is created.

A special result is the logfile, `ID.log`. It contains among various standard logging, the results of a number of plausibility checks (e.g. test for plausible river flows).

Another special results file is `catchment.thuban`, which is a session file for Thuban, a viewer for geographic information ([thuban.intevation.org](http://thuban.intevation.org)). GREAT-ER 2 is implemented ontop of Thuban.

### 4.1 Map Layers

- River Network Map Layer, Files `rivernet.(shp,shx,dbf)`
- Discharge Sites Map Layer, Files `discharges.(shp,shx,dbf)`
- Catchment Boundary Map Layer, Files `catchbound.(shp,shx,dbf)`
- Discharge-River Connection Map Layer, Files `disch_river.(shp,shx,dbf)`
- Picture Sites Map Layer, Files `pics.(shp,shx,dbf)`

## 4.2 River Network Attributes, File river.dbf

Fieldname	Type	Description
StretchID	I(10)	ID of the stretch (unique)
type	I(2)	-2, 0, 1 or 2 for OUT-FORMAT == 1.0. For OUT-FORMAT == 2.0, the type 6,8,9 and 10 (ie. base type +8) are the corresponding types indicating lakes.
IsDisch	I(1)	1 if a discharge site is connected to this stretch. Else 0.
up1	I(10)	StretchID of upstream stretch (corresponds to type)
up2	I(10)	StretchID of upstream stretch (corresponds to type)
down1	I(10)	StretchID of downstream stretch. Only for OUT-FORMAT == 2.0
down2	I(10)	StretchID of downstream stretch. Only for OUT-FORMAT == 2.0
rivclass	I(5)	ID for river class.
QMean	R(12,5)	Mean flow volume (assuming log-normal distribution) in $m^2/s$ .
Q5	R(12,5)	5th percentile of flow volume (assuming log-normal distribution) in $m^2/s$ . Sometimes also referred to as "95 percentile low flow"
vMean	R(12,5)	Mean flow velocity (assuming log-normal distribution) in $m/s$ .
v5	R(12,5)	5th percentile of flow velocity (assuming log-normal distribution) in $m/s$ .
RealLength	R(20,10)	The length of the stretch in $m$ .
depthMean	R(12,5)	Mean depth (assuming log-normal distribution) in $m$ .
depth5	R(12,5)	5th percentile of depth (assuming log-normal distribution) in $m$ .
Name	C(60)	A name for the stretch (river name).

## 4.3 Discharge Site Data, File disch.dbf

Fieldname	Type	Description
DischID	I(10)	ID of the discharge site (unique).
StretchID	I(10)	ID of the stretch the discharge is connected to.
MarketID	I(10)	ID of the Market information connected to this discharge.
Pop	I(10)	Population connected to the discharge (real population, not population equivalents).
DWF	R(20,10)	Dry Whether Flow in $m^3/s$ .
flow_dom	R(20,10)	Domestic flow in $m^3/s$ .
flow_nondo	R(20,10)	Non-domestic flow in $m^3/s$ .
flow_runof	R(20,10)	Runoff-flow in $m^3/s$ .
Flow	R(20,10)	Actual effluent flow in $m^3/s$
treated	R(10,5)	Fraction of flow that is treated.
WWTPID	I(10)	ID of the WWTP class of this discharge.
wwtp_frac	R(10,5)	Fraction that is removed by the WWTP.
Name	C(40)	Name of the discharge.

## 4.4 Discharge Classes, File wwtp.dbf

This file is created as a constant file by the pre-processing. For GREAT-ER 2, the classes are already incorporated into the database and will likely not require an update.

For units and further description, please refer to the corresponding documentation of GREAT-ER 1 and GREAT-ER 2, respectively.

Fieldname	Type	Description
WWTPID	I(4)	ID of the WWTP (unique).
mode	I(1)	(only GREAT-ER 1.X) default mode.
WWTPtype	I(1)	Sewage treatment processing type.
prim_pres	I(1)	(only GREAT-ER 1.X) Is a primary settler present (1), else (0).
cap_prim	R(10,5)	(only GREAT-ER 1.X) capacity primary settler.
depth_prim	R(10,5)	Primary settler depth.
v_over_pri	R(10,5)	Primary settler overflow rate.
R_SS_prim	R(10,5)	Maximum primary settler solids removal.
depth_sec	R(10,5)	Second settler depth.
v_over_sec	R(10,5)	Second settler overflow rate.
R_SS_sec	R(10,5)	Maximum second settler solids removal.
capa_sec	R(10,5)	(only GREAT-ER 1.X) capacity secondary settler.
nr_PE	I(10)	(only GREAT-ER 1.X) number of Population Equivalents.
SS_ML	R(10,5)	Mixed liquor SS.
recycle	R(10,5)	Sludge recycling ratio.
a_biomass	R(10,5)	Biomass activity correction.
depth_ML	R(10,5)	Mixed liquor tank depth.
B_X	R(10,5)	Sludge loading rate.
Y	R(10,5)	Biomass yield.
DO	R(10,5)	Actual DO in aeration tank.
aer_type	I(1)	Aeration type: 0=Bubble, 1=Surface.
f_aerobic	R(10,5)	Aerobic redox fraction in mixed liquor tank.
f_anoxic	R(10,5)	Anoxic redox fraction in mixed liquor tank.
f_anaerob	R(10,5)	Anaerobic redox fraction in mixed liquor tank.
eps_N	R(10,5)	Nitrification efficiency.
eps_DN	R(10,5)	Denitrification efficiency.
WWTPdescr	C(40)	Description of the WWTP class.

## 4.5 River Classes, File rivclass.dbf

This file is created as a constant file by the pre-processing. For GREAT-ER 2, the classes are already incorporated into the database and will likely not require an update.

For units and further description, please refer to the corresponding documentation of GREAT-ER 1 and GREAT-ER 2, respectively.

Fieldname	Type	Description
RivClassID	I(4)	ID of the river class (unique).
mode	I(1)	(only GREAT-ER 1.X) default mode.
IsLake	I(1)	(only GREAT-ER 1.X) is this river a lake?
cor_riv_cs	R(10,5)	(only GREAT-ER 1.X)
SS_mean	R(10,5)	Mean river suspended solids concentration.
SS_stdev	R(10,5)	Standard deviation river suspended solids concentration.
k_SS_mean	R(10,5)	(only GREAT-ER 1.X)
k_SS_stdev	R(10,5)	(only GREAT-ER 1.X)
f_oc	R(10,5)	River suspended solids organic carbon fraction.
v_sed_grow	R(10,5)	Sediment growth velocity.
eps_sed	R(10,5)	Sediment porosity.
ro_SS_sed	R(10,5)	Sediments solid particle density.
X	R(10,5)	Lumped river biomass level.
DO_mean	R(10,5)	Mean Dissolved Oxygen.
DO_stdev	R(10,5)	Standard deviation Dissolved Oxygen.
BOD_mean	R(10,5)	(only GREAT-ER 1.X)
BOD_stdev	R(10,5)	(only GREAT-ER 1.X)
pH_mean	R(10,5)	Mean acidity value.
pH_stdev	R(10,5)	Standard deviation acidity value.
a_rivclass	R(10,5)	(only GREAT-ER 1.X)
k_light	R(10,5)	Light extinction coefficient.

## 4.6 Background Data List, File backgroundlst.dbf

Only for OUT-FORMAT == 1.0.

Fieldname	Type	Description
Name	C[30]	Title of the layer.
Mapname	C[25]	Filename of the layer.
Type	C[20]	Type (e.g. polygon, image, grid).
Legendfile	C[25]	Filename of legend file.
ctch_fl	C[8]	"yes" or "no" indicating whether there is some special treatment (eg. for DCW).

## 5 Preprocessing Basics

Besides the environmental model itself, the availability and quality of environmental data mainly determine the validity of simulation results. This and the following sections introduce general aspects and problems of spatial data for environmental exposure assessment. A proposed procedure is described for how to deal with the collection, processing and quality assessment of data considering also the origin and purpose of readily available data. The technical aspects will be focused on but it must also be stressed that obtaining spatial data in the EU reveals serious coordination and authorization problems.

An example of the proposed procedure is given in which the digital river network is the backbone of the spatial data set for "down-the-drain" exposure assessment. The example is given for GREAT-ER but the methodology is equally applicable for similar systems.

The data required for the regional geo-referenced exposure assessment is characterized by its complexity and inhomogeneity. This is partly due to the fact that there are large amounts of single data. But another fact plays an important role: the composition of such data is not readily available. It is typically scattered among several competent authorities and other data owners. In the past those different parties had seldom experienced the necessity to harmonize their data and hence it is to be expected that all obtainable data are in several digital formats and/or geographical resolutions. Inconsistent ID reference systems are likely to be faced as well. The approach to regional geo-referenced exposure assessment has created a need for the coordination and composition of base data which are, though collected and applied for various topics, nothing more than *raw* data in respect to the new task.

The reason for this lies also in the way in which most geographic data are used: for mapping purposes where structural integrity is much less important than visual appearance.

The data composition for exposure assessment in river networks reveals many problem-specific aspects. When one would set up a very special recipe for certain data sources or data origins, this task would lack re-use capabilities and also would not be considered adequate for other comparable tasks. In the framework of a simulation system for a growing number of catchments, re-use capabilities require special attention to ensure the efficient creation of many digital river network data sets. The solution for data composition must be simple and open enough to offer quick understanding, general applicability and also to enable technical solutions to be adapted.

The usual steps of data composition and processing as a GIS task are typically single manual steps of applying geographic methods and routines. The required quality and structure of the final geographic data (i.e. maps) is approached in a straight-forward manner. The data sources and intermediate stages are often eliminated as soon as the latest version passes a quality control. It

is unusual for GIS users to automate the processing steps because each new job reveals other aspects of processing which can not all be foreseen. To make the results transparent and to avoid later wrong use, all processing steps must be documented very carefully. This is especially important when results are not only used for reporting purposes but also for further processing and comparison with other data sets processed in a similar way.

A well defined intermediate stage of the data is a general approach to separate the specific problems of raw data from the actual data preparation. Setting up this intermediate state on text-based formats increases transparency. A set of pre-defined files requests data at an adequate level of simplicity while at the same time being able to grant some aspects of quality and completeness. Furthermore, a high degree of automation can be created starting from pre-defined text files.

As an alternative intermediate state a pre-defined database management system (DBMS) can also serve the task of separation. Such an approach creates further needs in software requirements and skills to handle them. Depending on the frequency of usage, the DBMS solution, once set up and tested, can be more efficient than the solution based on simple text files. The advantages arise from the DBMS environment which offers support for distributed systems (including Internet applications) and high performance in handling large data sets. With its higher level of sophistication such a solution loses self-explaining transparency which needs to be compensated by comprehensive usage documentation and/or by dedicating a human operator. Proving its principle operability with the text files as an intermediate stage is a recommended step before setting up a DBMS solution. The two-step approach is therefore chosen for the GREAT-ER data processing.

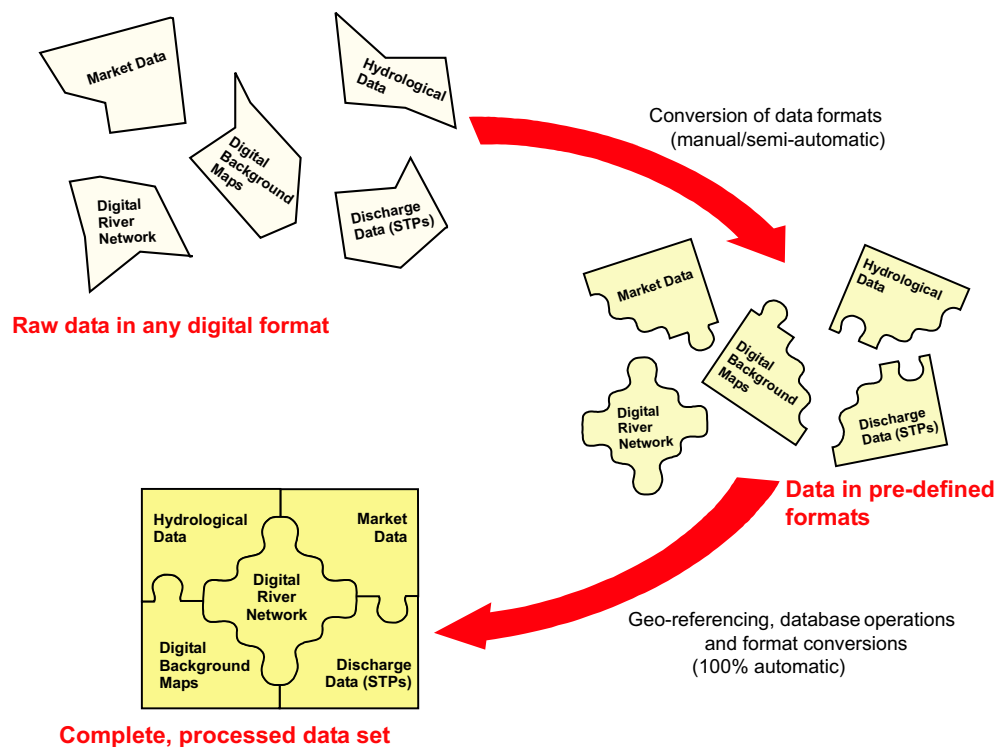


Figure 5.1: *Two steps of data processing*

Inventing an intermediate stage for data consequently results in two steps of data processing. In a first step, collected raw data are transformed into the pre-defined intermediate form. A second and independent step then creates final ready-to-use data from the intermediate stage.

The initial data processing is very much determined by the gap between demanded data quality and formats and the present state of the raw data. The required files have to be produced manually from the original raw data or with special routines which probably have to be developed anew.

The final data processing is a well defined task, provided that all required data are collected and fit the demands of pre-defined formats. Of course, the final environment of the geo-referenced exposure assessment tool must also be defined precisely to make the data processing fully automatable. In this context, the simulation method determines the structural shape and order of the data while the user interface and data management modules determine the actual formats of files.

The two steps of data processing are illustrated in figure 5.1. An example is given for an exposure assessment tool for river networks. Three main states of the five data groups are shown. These data groups cover any information that is related to space in one way or another. For example, physico-chemical data of a substance are excluded but the substance's market data are included as regionally resolved information.

Another view of the finally produced data sets is shown in figure 5.2. The data groups form layers which build the visual interface available to the user.

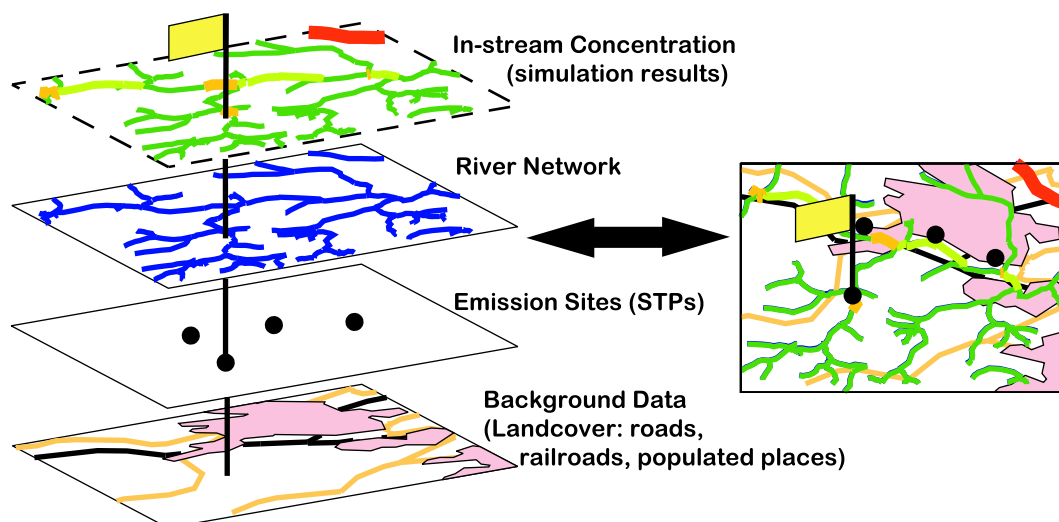


Figure 5.2: *Layered view of data groups*

## 5.1 Data requirements and composition

### 5.1.1 Digital river network

#### General overview

The digital river network describes the structure of the river network in question. At the very bottom it consists of lines (also referred to as 'polylines' or 'arcs'), each of which represents a river segment. Furthermore, it consists of nodes which represent the point of connection of two segments. Lines are expressed as a sequence of at least two coordinate pairs, nodes are expressed as exactly one coordinate pair (figure 5.3b).

Lines have to be considered as vectors; the first point of the sequence is the starting point and the last is the end point. The vector direction is used to represent the flow direction (figure 5.3c). The length of a segment can be computed from the coordinates.

Each transition from one segment to a subsequent one (sequence, confluence or bifurcation) is defined by a node which contains the information 'from' and 'to'. The ID's of the upstream ('from') and the downstream ('to') segments are stored. The geographic position of a node is uniquely defined with the last point of the upstream and the first point of the downstream segment. If these two points are not identical, the two stretches are not connected and no corresponding node exists.

Besides the persistent Arc-Node geometry model, a second, flexible segmentation can be set on top of it. The two-dimensional geometry model lacks the ability to precisely locate sites along the lines between two given coordinates. This can be solved by introducing a one-dimensional measuring system with only starting and ending information.

These one-dimensional geometric objects are called 'routes'. For digital river networks, each river is a 'route'; the whole network constitutes a 'route system'. Any site along the rivers in a route system can be referenced by the route name and the distance from the route start (e.g. River Rhine at kilometer 744 which is at Düsseldorf, Germany). The start and end of routes are a matter of definition. In Germany, the distance reference of rivers that lead into the ocean begin at the source. The distance reference of tributaries begins at the point of confluence in reverse flow direction. Standard GI Systems offer methods to create route systems from an arc-node model. This must be a two-step routine where a raw route system is first created based only on the information contained in the arc-node representation (figure 5.3d). The raw route system requires two further adjustments. First, the digitized rivers do not necessarily start at the actual source and hence the starting points of the routes must be set to the correct value for later consistent referencing. Second, the length computed from the digitized lines depends on the resolution and hence is always lower than the actual river length. A remeasurement routine that is provided with the actual length of a route, is to be performed to gain revised coordinates for any event along the stream. The route system approach for exposure assessment was applied for intermediates in the River Rhine catchment (Koormann, 1997 and Koorman et al. 1998) and for a small tributary of the Rhine, the Itter catchment (Schulze, 1998 and Schulze et al., 1999).

Fundamental to the creation of a digital river network are maps of the corresponding area. Traditional printed maps can be used as a basis but usually such maps now exist in digital format.

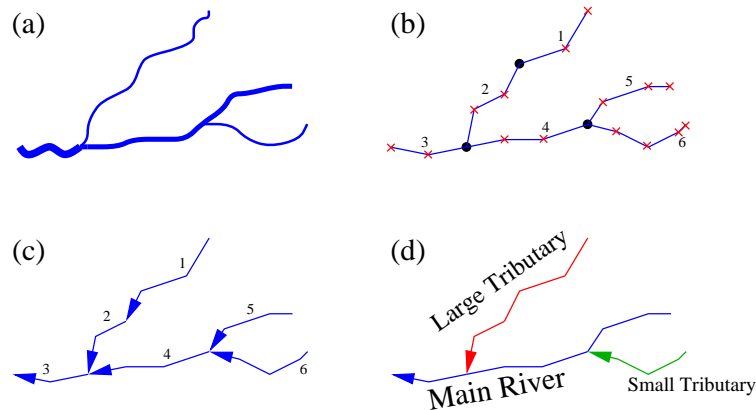


Figure 5.3: *Steps of creating a digital river network: (a) Reality (i.e. aerial photo or comparable basis). (b) Digitized network (crosses and dots = coordinates, dots = nodes, numbers = IDs of lines). (c) Direction of lines (vectors). (d) Representation as route system.*

Digital maps can be created in two ways, either by collecting coordinates directly with a Global Positioning System (GPS) or from aerial photographs. While GPS data are naturally already geographic objects in vector format, pictures taken from planes or satellites offer information in a raster format. I.e., surface waters consist of single points (squares) and hence do not contain information about e.g. flow direction, but on the other hand may offer information about areal aspects (e.g. width of rivers, shape and size of lakes) and possibly further properties (temperature, depth). The raster information must be converted to vector data to gain a digital river network as described above. This conversion can partially be automated but manual reworking by human experts is unavoidable for quality assurance.

### Solution for GREAT-ER

As a requirement for data, the arc-node model is preferred rather than the route model. This is because, especially for more detailed river networks, an assignment of river names and remeasurement become problematic and time-consuming.

In summary, the following demands for a digital river network have been defined:

- Only the coordinate sequences for the river segments are to be given. No extra node specification is required.
- Should two segments be connected, it is necessary that the end point of the upstream and the starting point of the downstream segment are identical. This also means that for confluences or bifurcations the upper segment(s) always end at that point and one (or two) new segments start from the same point.
- All river segments must be directed downstream.
- All segments must have a unique ID.

Only confluences and bifurcations enforce the beginning of new stretches. Sequences like segments 1 and 2 in figure 5.3 have no topological meaning for the network. Splitting up segments into a sequence of two or more stretches becomes important for the semantic purposes the digital river network has to serve. Simulation models using digital river networks may need to simplify river segments by assuming stretches to be homogeneous in a hydrological sense. In this case flow data would be attributed to river stretches and hence long segments need to be split up according to the required resolution or available knowledge about hydrological changes. Additionally, for visualization purposes the level of segmentation also determines the capabilities to graphically display stretch-specific information. Hence, when setting up the digital river network the later purpose of these data should already be considered at this point and segmentation should be performed. This task can also be undertaken automatically provided that the single coordinates of the lines are more or less equidistant.

### 5.1.2 River network attributes

Depending on the needs of an applied exposure simulation model, information about certain properties of the river network are required. Usually these are floating point numbers (e.g. flow in [ $m^3/s$ ] or flow velocity in [ $m/s$ ]). Further attributes such as boolean (shipable [yes|no]) and others are conceivable. Most of the traditionally measured data are expressed as statistics due to natural fluctuations and seasonal variation. For example, flow data are usually expressed as log-normal distributions characterised by the mean and standard deviation (or mean and low, e.g. 5th, percentile).

However, all discrete attributes (in space) have in common that they correspond with a certain river segment. A set of discrete parameters for a river network can hence be given in tabular style. Continuous parameters (in space) are more difficult to attach to digital river networks, and applications using such data also reach a level of complexity which is seldom acceptable for actual environmental simulations. Usually distribution curves are derived from temporal data series which are then applied for a specific task.

Depending on the purpose for which the river attributes are needed, in some cases the variation of each river segment is of importance and for other parameters variations are only of interest for entire rivers or river networks. Furthermore, decisions regarding the stretch-wise resolution of parameters are influenced by data availability. Consequently, a task-specific separation of parameters is helpful and can be solved by defining river classes and associating each river segment to a river class.

Parameters that are semantically important and presumably different for each segment are directly associated to river segments. For many parameters the reasons for actual availability at high resolution represents the opportunity not only to measure but also to interpolate between measurements or to estimate them from other attributes. Hence, again depending on the semantic context, not all parameters are obligatory. A flexible (interpolating, estimating) system of optional parameter specification is possible.

The methods of chemical exposure assessment are organized in a tiered approach in terms of physico-chemical processes. Lumped transport and elimination processes will be applied first and if it is not capable of explaining monitored behavior, refined substance-specific processes will be

added (e.g. photodegradation). Any refinement increases data demands, uncertainties and, on a very practical side, expense and time.

The counterpart of the lumped parameters for chemicals (transport and elimination) are hydraulics and geometry of rivers. These basically determine the dilution and dispersion of chemical loads. Accompanied by length and name, these were chosen as direct attributes. All other parameters are assembled in river classes. This choice reflects the data requirement of the simulation software incorporated in GREAT-ER (Boeijs, 1999). The concept of the intermediate data stage still allows us to consider data requirements of any comparable simulation model by providing corresponding automatic data processing routines for the second step (figure 5.1).

## Flow, Velocity, Depth and Width

**Overview of interrelation** For a sufficient period of time, the flow passing the full extent of a river segment can be assumed as constant in space and time. In this case the continuity equation 5.1 is true.

$$Q = vA \quad (5.1)$$

where  $Q$  = flow,  $v$  = velocity and  $A$  = cross-sectional area.

Rivers have no rectangular cross-section and hence  $A$  can not simply be substituted by the product of depth and width. Depth and width have a correlation with the flow within certain ranges. A set of power equations 5.2 can express the interrelation of velocity ( $v$ ), width ( $w$ ), depth ( $d$ ) and flow (Leopold & Maddock, 1953).

$$v = aQ^\alpha \quad ; \quad d = bQ^\beta \quad ; \quad w = cQ^\gamma \quad (5.2)$$

where both  $\alpha + \beta + \gamma$  and  $a \cdot b \cdot c$  must be equal to 1.

Compared to velocity and depth, width is generally less variable (in space) for most rivers (Chapra, 1997). Especially when dealing with one-dimensional models, width is of less semantic importance. In contrast, depth is important, i.e. for chemical volatilization processes. In this approach width is hence dropped from the list of required river properties. It should of course be taken note of when considering transversal dispersion (two-dimensional river models).

Site-specific measurements and site-specific regression functions are optimal to gain most adequate information on river properties. In the absence of such data, and this is true for most sites, literature or new field studies need to be undertaken.

Numerous studies were conducted to obtain values for parameters  $a, b, c$  and  $\alpha, \beta, \gamma$  which vary with different climatic location and degree of artificial influences on the river networks.

It has also been shown that the hydraulic geometry can alternatively be obtained from the corresponding watershed size: Based on a strong correlation of bankfull flow and watershed area, a geometric relationship was shown and extrapolated for all river segments (Miller et al., 1997). Another method is based on the hortonian stream order for river networks. This ordering method attaches the order of 1 to all source stretches. The order of any other stretch is the maximum order

of the directly connected upstream stretches plus 1. A good correlation of the hortonian order and the river geometry has been proved for a watershed in Arizona, U.S.A. (Miller et al., 1996).

Besides the statistical approaches described above, mechanistic techniques can be applied. The best known method is the Manning equation (5.3) (Chapra, 1997).

$$v = \frac{R^{2/3} S^{1/2}}{n} \quad (5.3)$$

where  $n$  = Manning's roughness coefficient [-]  
 $R$  = hydraulic radius (equal to  $A/P$ ) [m]  
 $P$  = wetted perimeter [m]  
 $S$  = energy slope (here equal to geological slope) [-]

Substitution into the continuity equation results in the approach of Manning-Strickler (equation 5.4) (Dyck & Peschke, 1989).

$$Q = \frac{AR^{2/3} S^{1/2}}{n} \quad (5.4)$$

Besides flow and velocity, the depth can also be obtained based on the Manning equation. The depth is the root of equation 5.5 (Chapra, 1997).

$$f(y) = \frac{1}{n} \frac{[(w_0 + sy)y]^{5/3}}{(w_0 + 2y\sqrt{s^2 + 1})^{2/3}} S^{1/2} - Q \quad (5.5)$$

where  $s$  = side slope and  $w_0$  = bottom width.

For the GREAT-ER project a new study has been undertaken to estimate the velocity of any river reach in the UK (Round et al., 1998). A statistical approach was chosen based on hydrological databases for the UK. The quality of the velocity estimation model (equation 5.6) was not significantly reduced in terms of performance by excluding the hydraulic radius. The error analysis showed that with 68% confidence the true velocity lies within a factor of 1.59.

$$v = 10^{-0.583} Q_{mean}^{0.283} \left( \frac{Q}{Q_{mean}} \right)^{0.495} \quad (5.6)$$

To obtain average depth values, variations in three dimensions have to be made. First, depth can vary in transversal and longitudinal extent of the river bed and next, it changes in time due to flow variation. For unknown geometry and known flow an application of the Regime Theory provides equations 5.7 and 5.8 (Simons & Albertson, 1960).

$$w = 6.175\sqrt{Q} + 0.305 \quad (5.7)$$

$$d = \begin{cases} 0.610 + 0.93r & : r > 2.13 \\ 1.21r & : r \leq 2.13 \end{cases} \quad (5.8)$$

where  $r = 0.57Q^{0.36}$ .

**Selection of values and methods** Literature review has shown examples of the ability to estimate the interrelated river properties flow, velocity, depth and width from each other in several ways. Also area and mathematical topology can be included for estimations. The latter can even be applied when no stretch specific properties are known.

In general, higher efforts are required to measure the geometry of river segments while flows are easier to obtain: In most European countries a system of gauging stations is in operation, delivering flow statistics for some sites. Usually interpolation equations can easily be set up or are even readily available for ungauged sites between two gauging stations. Ungauged headwaters are more complicated to establish flow estimation. The flows can be computed from geo-morphology and other information, i.e. MicroLowFlows additionally applies soil type and precipitation information for this task (Gustard et al., 1992).

For the methodology introduced in this chapter, flow data are chosen as obligatory, and velocity and depth are optional parameters. Optional parameters will be estimated during the data processing. Consistency checks based on the literature can be integrated in the final processing.

Most estimation equations derived from statistical studies are associated with several limitations concerning the environmental framework of the investigated river network (natural/regulated, arid/humid, lowland/highland, etc.). If estimation equations are available which were set up for the same or a similar type of river these should be preferred and the resulting values should be used. Automatic selection of estimation routines can not decide adequately compared to human judgement. This is due to the fact that much general information on a river network is required and this information can not easily be formalized.

One important property of the discussed parameters is variation in time, which means that these parameters usually consist of three subparameters describing their statistical distribution (i.e. distribution type, mean, standard deviation). This must be kept in mind when dealing with corresponding data collection and estimation.

## Length

In conjunction with velocity, the length basically determines the transport of a (chemical) load. The process of transport in terms of chemical exposure modeling is a relatively fast process and hence the travel distance should be given as accurately (feasibly) as possible.

Obtaining the actual length of rivers should in general not pose a serious problem because it is relatively easy to measure. For some countries, canoeing guides offer this information in great detail, i.e. for Germany (Deutscher Kanu-Verband, 1985).

In theory, there are of course no problems in calculating the length of each segment based on coordinates. These values are naturally (sometimes much) lower than the real length of the river segments. These differences become obvious when taking into account that even detailed digitizing (e.g. one point every 300m) can not reflect the actual shape of a very curvy river. Additionally, the digitizing itself reveals uncertainties.

## Name

Names for each river segment are not meant to identify stretches in terms of database operations. Obviously one river can be split up in very high numbers of stretches and therefore unique name specification does not make sense. The names are required rather for identification and orientation purposes in the final simulation tool. Names of rivers can hence be associated with other site names or any information that is felt to be helpful to users.

## River Classes

The composition of the river class parameters is to be selected dependent on the semantic purpose of the simulation models. This way, the number of properties can be reduced to a minimum to retain low complexity and an easy overview.

The methodology presented in this chapter takes the river class composition as used for the simulation software in GREAT-ER (Boeije, 1999). Included are typical water quality parameters, i.e. suspended solids, dissolved oxygen, biological oxygen demand and pH. All these parameters are only used in higher model complexity modes, i.e. not applying lumped degradation rates. Among the biochemical properties, the type of segment is also described (river or lake).

### 5.1.3 Catchment boundary

The catchment of a river network is defined by the most downstream point, i.e. the most downstream segment. The catchment boundary represents the natural drainage area. Any point in that area will drain to that most downstream segment. The catchment boundary must always be a closed polygon.

The introduced data composition additionally enables several separated river networks to be joined. This helps users to deal with e.g. flat coastal regions, entire islands or even whole countries. In such cases the catchment boundary should be the union of all single subcatchments.

Information about the actual border of a catchment serves two purposes. First, it supports geographic methods, e.g. spatial selection routines. Second, within the final tool the user is given further support for orientation and interpretation.

In regions where water management has a long tradition, watershed boundaries should be readily available. Any new or re-calculation applying a GIS will require digital elevation data. This computation is often performed and several technologies are available, partly already built into a GIS.

### 5.1.4 Discharge data

The discharge data, in this context, are a set of points through which substances are released into the river networks, e.g. sewage treatment plants. The discharge information will not be able to express diffuse input (unless it is made discrete as a series of point inputs). Diffuse input is not focused in the presented methods because a different class of simulation models would be required.

Discharge points with geographical reference define an area where any release will pass this point travelling "down-the-drain". Such areas are scale-independent geographical units interconnected by a network topology (figure 5.3). This scale independency allows us to flexibly *refine* or *aggregate* (parts of) the geographic data where required. Refinement on the one hand offers dynamic, problem-oriented geographical resolution to avoid unnecessary data collection and helps to deal adequately with data gaps and to zoom into areas of special interest. Aggregation on the other hand enables us to reduce complexity, to compare regions by abstracting data and to calibrate actually needed scales of data.

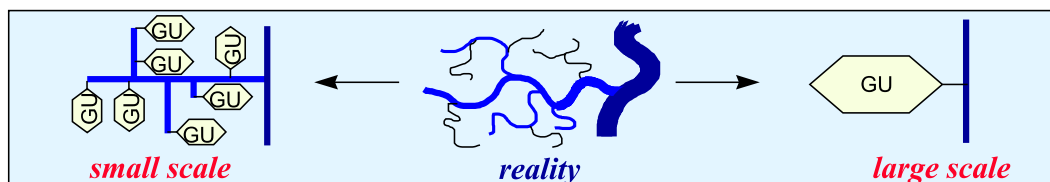


Figure 5.4: *Geographic units (GU): scale independent*

**Coordinates** The coordinates of a discharge site are obligatory. In the case of treatment works they describe the position of the actual plant and not the point at which the emission enters the river. The connection to the river will be established by explicit specification of the corresponding river segment. This enables provision of location information to the user who is supported in interpretation by offering the distance of the plant to the river and whether the emission takes place on the left or right bank.

**Population** For simulations of chemical ingredients in consumer products, the per capita consumption is usually available in mass per time. To properly estimate the release, the actual real population connected to a discharge is required. These numbers should be readily available from responsible environmental protection authorities, water authorities or water supply companies. For purely industrial discharges the population is zero.

It is not feasible to use population equivalents, because these numbers include industrial emissions based on a special selection and weighting of water quality parameters. For example, in Germany a per capita emission of 200 liters of waste water with an equivalent load of 60 g of biological oxygen demand (BOD) is assumed. Based on this, several population equivalent types are defined (DIN 4045 (12/1985); Römpp, 1997): Equivalents based on BOD (EGWB60), based on waster water emission (EGWW200), resulting sludge volume (EGWS2,0) and resulting dry sludge (EGWTS80). Across Europe several national standards are to be expected; for some countries variations between water authorities are even possible (e.g. in the U.K.). For most consumer substances no correlation with the standard water quality properties exists and hence population equivalents may reveal helpful background information, but are not required for chemical exposure assessment.

**Flows** Flows within the waste water and associated paths are measured and defined in several ways across Europe. For the data requirement definition a most common and practical selection is

presented here. Above all the most important value is the flow leaving the plant and entering the river network. The actual daily flow (ADF) will mean the average (measured) emission of a plant in  $m^3/d$ .

In many cases consented maximum emissions are available. Such values are only variable with the consenting interval and do not necessarily represent the real emissions. For those sites for which no ADF is available the consented daily flow (CDF) is the best approximation, since the operators of waste water treatment facilities usually try to fill the consented flow.

The inflow of a treatment plant consists of three components: Runoff, non-domestic and domestic flow.

**Discharge Type** The type of a discharge is considered in the simulation with specific models to estimate removal efficiency. Usually the technological type of the corresponding waste water treatment plant is required (i.e. trickling filter plant, activated sludge plant), but the discharge type can be "direct discharge" (no chemical removal) as well.

Some sewage treatment plants are actually a combination of different technologies. A specific modeling of such situations is not feasible. A conservative simplification choosing the less efficient type is recommended, but for transparency the situation should be noted (see the *Name* below).

**River connection** Discharges are connected to the first point of the given segment which is to be specified with its ID. For an optimal representation of the actual river entry point, the river network digitizing or segmentation should already consider these points.

In principle, GIS routines (distance measures) can be applied to find out the most reasonable stretches. These methods are automatic but always require human plausibility control.

**Name** The only purpose of the names is to support user orientation and interpretation as well as reporting. The names need not to be unique and should contain further information or hints when necessary.

### 5.1.5 Background maps

Though not used in the simulation models, background information is important for the final end user system. They allow visual orientation and plausibility control as well as aiding interpretation of simulation results, which is carried out by the user.

Different types of background information can be provided: raster landcover maps, vectorized information (geographic objects) on e.g. soil types and land use. Additionally, semantically related (temporal) information can be prepared such as monitoring sites with attached measured time series.

## 5.2 Final automatic data processing

The final data processing requires syntactically and semantically correct source files of the files ".drn", ".rna", ".dsd", ".cbp" and ".bgd" as described above. The whole processing with its subsequent steps is managed by the main program *greater-pre-processing.py*.

This part of the data processing is, in principle, not dependent on the required simulation system, though it does simplify the simulation system and user environment developed within the framework of GREAT-ER. The simple tabular structures allow the easy extension of the data files for adaptation to any other comparable simulation system.

The data to be generated can be divided into two groups: geographic data and the corresponding attribute data. The geographic data form the basis for visualization and geo-referenced data processing. These elements (cf. points, lines and polygons) are associated with attribute data via unique ID's. Their interconnection is based on the same ID reference.

Processing of the background data (".bgd" file) is not related with the main processing due to its special task to support the user. It is not discussed further in this document.

### 5.2.1 Geographic data

The establishment of geographic data consists of two processing steps (figure 5.5). First, the geographic elements (their ID's and coordinates) are extracted from the files in the pre-defined formats and converted into another format that can be handled directly by a GIS that handles the Shapefile format.

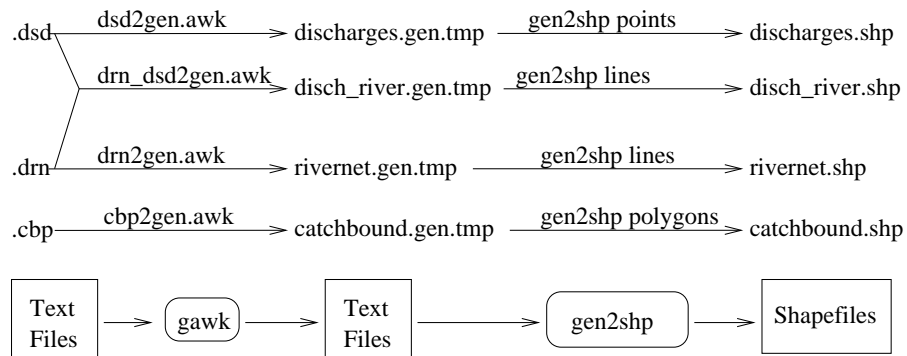


Figure 5.5: Command sequence for automatic processing of geographic data

A straightforward conversion is executed for the two digital maps: discharge sites (".discharges.shp") and river network (".rivernet.shp"). For the boundary of the catchment (".catchbound.shp") a further attribute is appended from the datafile "catchments.txt": the full name of the catchment.

The connection lines between discharges and corresponding river segments are computed using information from the ".dsd" and ".drn" files. A special shapefile is created (".disch\_river.shp"), which provides visual information on where the discharge is connected to the river network.

## 5.2.2 Attribute data

While the attribute data are being processed (figure 5.6), the three main data files (".dsd", ".rna" and ".drn") are first joined and then create the primary attribute files for river segments ("river.att") and discharges ("disch.att") as well as the secondary attribute files for treatment plant types ("wwtp.att") and river classes ("rivclass.att"). The two secondary files are (currently) static and basically describe generic treatment plants and rivers.

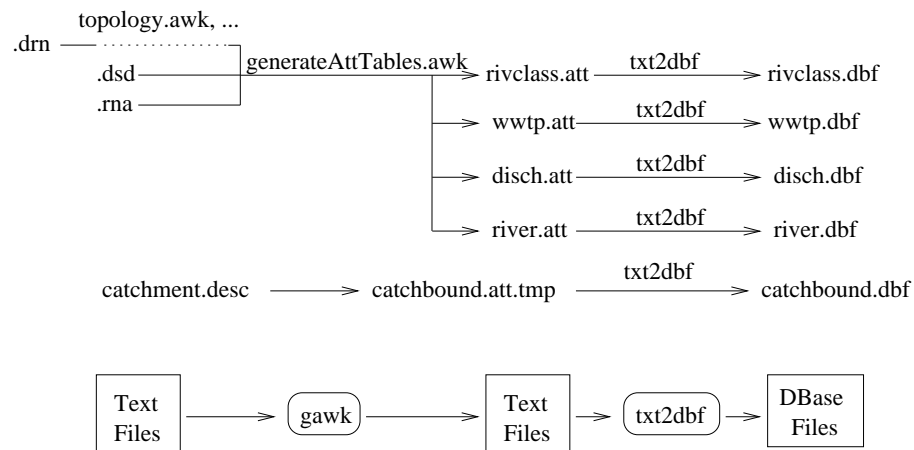


Figure 5.6: Command sequence for automatic preparation of attribute data

In the final processing step, text based tables are converted into DBase format. The purpose of this is to optimize data management within the visualization tool.

## 5.2.3 Core algorithms

Several data processing steps consist of simple routines (e.g. format conversions). They are not explicitly explained here. In this section the core problems of data processing are discussed.

### Creation of River Stretch Sequence

The information about the logic sequence of single river segments plays an essential role for simulation models. A simulation model must be able to *walk* down from any source to the final mouth. Hence, each stretch must offer information (the ID) about its successor. This purpose is served by a simple two-column table `fromStretchID, toStretchID`.

The ".drn" files, if semantically and syntactically created as described above, contain the topology information, but coded in the given coordinates. The arc-node model is used to convert the topology to the coding based on StretchIDs. Building the arc-node model from a ".drn" file associates all segments with IDs of the corresponding "from-node" and "to-node" to describe their direction. Consequently, the two-column tables `fnode, StretchID` and `tnode, StretchID` can be extracted directly from the arc-node model ("topology.awk"). Joining these two tables by the node ID and then omitting the node ID results in

the two-column table `fromStretchID, toStretchID`. In both columns IDs may now occur multiple times due to confluences and bifurcations. To create a table where the `fromStretchID` is a unique identifier, the `toStretchIDs` are joined up into a (space character divided) list. The resulting two-column table `fromStretchID, toStretchIDlist` is a straightforward representation of the downstream topology. Analogously, an upstream topology table `toStretchID, fromStretchIDlist` can be created starting with the `fromStretchID, toStretchID` table and exchanging the position of the two columns.

### Creation of Binary River Network Topology

Fast and straightforward access for any algorithm walking up or down the river network is realized using simple data tables. The minimum number of columns to describe any river network topology is four. Demanding one column being a unique segment ID and also allowing only exactly one ID in any field, virtual segments need to be added to grant topologic consistency (e.g. confluence of three streams at one point). An additional column finally describes the segment type (listed in table 3.1). This is needed to correctly interpret the content of the other columns. Extending the network by virtual segments will not increase but decrease the total table size because a further column would rarely be filled for natural networks where a confluence of two streams is usually the maximum. For artificial networks (e.g. sewer systems) the situation can be different.

Table 3.1 *Segment Types*

Type	Type (if segment is a lake; only GREAT-ER 2.0)	Number of segments		Description
		Upstream	Downstream	
0	8	0	0,1,2	Source
1	9	1	1,2	Normal segment
2	10	2	0,1	Confluence
-2	6	1*	0,1,2	Bifurcation

\* which must be of type 1 with 2 downstream segments

Algorithms for walking on binary trees from the leaves to the root (cf. downstream) are typically implemented recursively starting with the root (cf. mouth). For exposure assessment in river networks the binary recursion has to be "post-order" which means performing an action for a segment after all upstream segments have already undertaken this action. Hence, the upstream direction is required for the two-column topology table.

The above-described tables for the river stretch sequence are the basis for building the binary topology. During the data loading routine, the highest segment ID is looked up. This is important to invent further ID's for virtual segments.

When all data are loaded, first bifurcation situations are searched for. The corresponding segments (type "-2") are associated with their neighboring segments. This is important in order to reduce the final topology processing to a one-pass computation (a single linear processing step). Next, the segment table is processed straightforwardly. Depending on the number of upstream and downstream segments, the type is then decided. Several special cases need further treatment i.e. insertion of virtual segments:

**(I) Confluence of three segments at one point** When exactly three segments are directed to the same point, a new (virtual) segment is inserted as shown in figure 5.7.

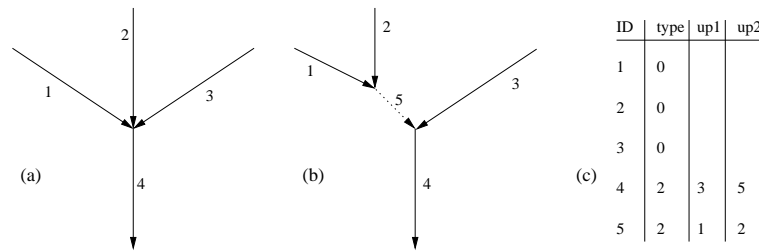


Figure 5.7: Resolving confluence of three segments (a) into binary representation (b). Numbers are stretch IDs. The table of binary network topology of (b) is (c).

The virtual segment obtains a length of 0. The ID does not appear in the geographic data of the river network and is neutral for the database operation "join". The simulation software must of course know how to treat segments with a length of 0: The results (e.g. chemical loads) must be transferred one-to-one to the next downstream segment.

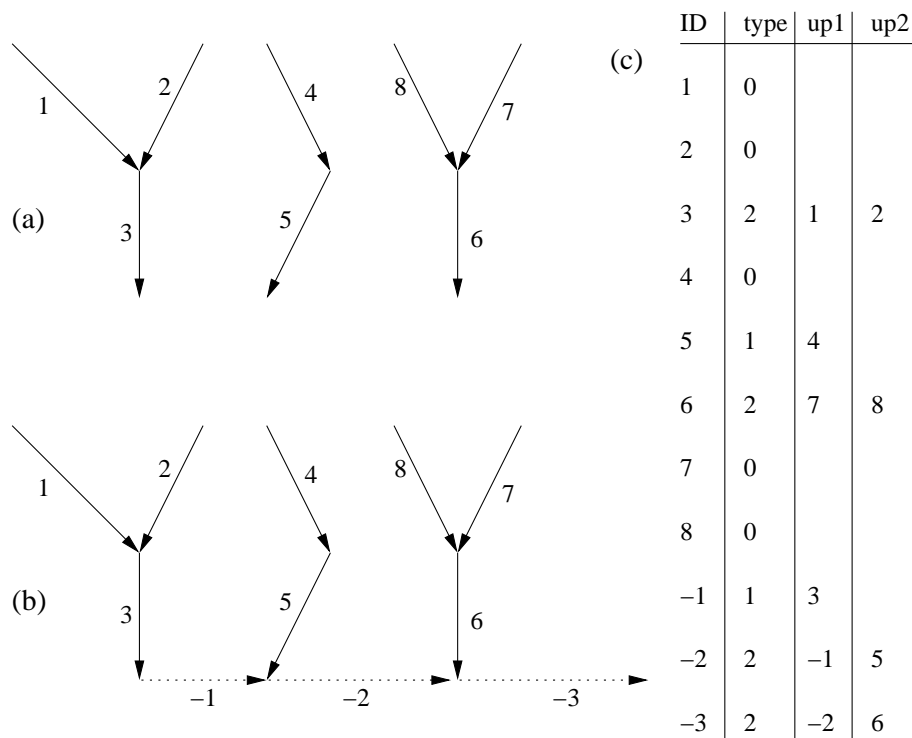


Figure 5.8: Resolving multiple end segments (multiple catchments) (a) into one catchment (b). Numbers are stretch IDs. Table of binary network topology of (b) is (c).

**(II) Several end segments** If several catchments are joined (e.g. to cover an entire island) the problem of multiple start segments (mouths) occurs and needs to be incorporated into the topology. The solution is a linear interconnection with virtual segments. By definition all "ocean segments" are attached with negative ID's.

In the example shown in figure 5.8, the lowest ID is -3 and it is the new starting segment for the resulting catchment. The ocean segments influence neither visualization nor simulation.

**(III) Bifurcation** For bifurcation situations both neighbours are of the type "Bifurcation" (figure 5.9). Field up1 will contain the ID of the upstream segment and field up2 will contain the ID of the neighboring segment.

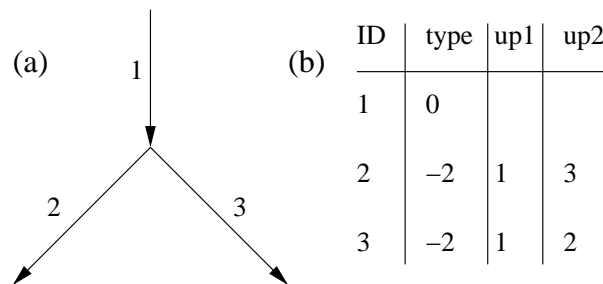


Figure 5.9: Handling of bifurcations (a) in the binary presentation (b).

**(IV) Unconsidered situations** Some situations may occur in the digital river network, but are unlikely to happen in reality for natural catchments. In practice it is more likely that wrong digitizing or another error may cause the situation. Such situations are not considered in the presented methodology though it is possible to resolve them with further virtual segments or additional definitions.

**(IV.a) Multiple confluence** More than three segments are directed to the same point (figure 5.10a). This situation could be managed similar to the confluence of three segments using virtual segments, but from a hydrologic point of view this situation is most unlikely and the chance of using wrong base data is much more probable.

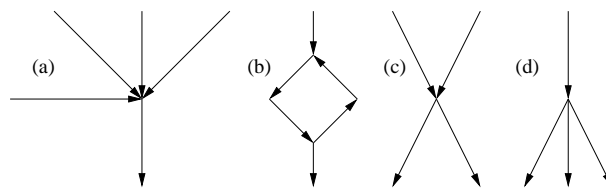


Figure 5.10: Unconsidered situations: Multiple confluence (a), circle (b), crossing (c) and multiple bifurcation (d).

**(IV.b) Crossing streams** Both, a confluence and a bifurcation take place at one point (figure 5.10c). Such a crossing is unlikely. Furthermore, a chemical fate model would need additional

information to adequately deal with such a situation (what fraction of which upstream segment goes into which downstream segment?).

**(IV.c) Circular flow** The segment sequence forms a ring (figure 5.10b). Building a ring, a river network is in topological terms not a directed graph. The applied simulation system is not capable of dealing with this situation. Such circles may occur, for example, in estuaries where the flow direction changes over time following a certain probability distribution.

**(IV.d) Multiple bifurcation** Flow separation of one segment into more than two segments (figure 5.10d). This situation is again unlikely from a hydrological point of view. An error in the base data is to be supposed.

## References

- Boeije, G. (1999). Chemical Fate Prediction for Use in Geo-Referenced Environmental Exposure Assessment. PhD thesis, University of Ghent, Belgium.
- Chapra, Steven C. (1997). Surface Water-Quality Modeling. McGraw-Hill, New York.
- Deutscher Kanu-Verband (1985). Deutsches Flußwanderbuch, 22. Auflage, DKV Wirtschafts- und Verlags GmbH, Duisburg.
- Dyck, S. & Peschke, G. (1989). Grundlagen der Hydrologie. Verlag für Bauwesen, Berlin.
- Gustard, A., Bullock, A. & Dixon, J.M. (1992). Low flow estimation in the United Kingdom. Institute of Hydrology report 108, Wallingford.
- Koorman, F. (1997). Organische Schadstoffe im Fließgewässernetz Rhein Modellierung, Simulation, Visualisierung von chemischen Zwischenprodukten. Diplomarbeit, University of Osnabrück.
- Koormann, F., Matthies, M., Trapp, S. (1998). Simulation and Visualisation of Spatial Exposure Patterns: Intermediates in the Rhine River System, In: Regulation for Chemical Safety in Europe: Analysis, Comment and Criticism (Eds. D.M. Pugh, J.V. Tarazona), Kluwer Academic Publishers, Dordrecht, pp. 65 - 96.
- Leopold, L.B. & Maddock, T. (1953). The Hydraulic Geometry Channels and Some Physiographic Implications. Geographical Survey Professional Paper 252, Washington DC, USA.
- Miller, S.N., D.P. Guertin and D.C. Goodrich (1996). Linking GIS and Geomorphologic Field Research at Walnut Gulch Experimental Watershed. Proceedings of the AWRA's 32nd Annual Conference and Symposium: "GIS and Water Resources", Sept. 22-26, 1996, Fort Lauderdale, FL. pp. 327-335.
- Miller, S.N., M. Hernandez and L.J. Lane (1997). GIS applications in the spatial extrapolation of hydrologic data from experimental watersheds. Proceedings of the Conference on Management of Landscapes Disturbed by Channel Incision, Oxford, MS, USA, May 20-22, 1997.
- Römpp Lexikon Umwelt (1997). Version 1.0, Georg Thieme Verlag, Stuttgart / New York.
- Round, C.E., Young, A.R. & Fox, K. (1998). A Regionally Applicable Model for Estimating Flow Velocity at Ungauged River Sites in the UK - Contribution to GREAT-ER #4. Water and Environmental Management 12, pp. 402-405.

- Schulze, C. (1998). Modellierung und Visualisierung der Konzentration von Waschmittelinhaltsstoffen in der Itter. Diplomarbeit, University of Osnabrück.
- Schulze, C., Matthies, M., Trapp, S., Schröder, F.R. (1999). Georeferenced Fate Modelling of LAS in the Itter Stream. *Chemosphere*, 39(11), 1833-1852.
- Simons, D. B. & Albertson, M. L. (1960). Uniform water conveyance channels in alluvial material. *Proc. Am. Soc. Civ. Engg., J. Hydraulics Div.*, 86, pp.33ff.

# A GNU Free Documentation License

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.  
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

## PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format,  $\LaTeX$  input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by refer-

ence in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections A and A above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- State on the Title page the name of the publisher of the Modified Version, as the publisher.
- Preserve all the copyright notices of the Document.
- Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- Include an unaltered copy of this License.
- Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

- For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.
- Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.
- Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles. You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section A above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else

a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

## **COLLECTIONS OF DOCUMENTS**

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## **AGGREGATION WITH INDEPENDENT WORKS**

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## **TRANSLATION**

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section A) to Preserve its Title (section A) will typically require changing the actual title.

## TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

## ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.