

Eurowaternet: towards an index of quality of the national data in Waterbase

Prepared by:
Tim Lack, Michael Gardner and Ruth Cullingford
WRc

Project Manager:
Anita Künitzer
European Environment Agency



Layout: Brandenburg a/s

Any queries on this report should be addressed to the authors at: European Topic Centre on Water, WRC, Blagrove, Frankland Road, Swindon SN5 8YF, Wiltshire, United Kingdom.
Telephone: (44-1793) 86 50 00, www: <http://water.eionet.eu.int>

Legal notice

The contents of this report do not necessarily reflect the official opinion of the European Commission or other European Communities institutions. Neither the European Environment Agency nor any person or company acting on behalf of the Agency is responsible for the use that may be made of the information contained in this report.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (<http://europa.eu.int>)

©EEA, Copenhagen, 2003

Reproduction is authorised provided the source is acknowledged.

ISBN: 92-9167-578-4

European Environment Agency
Kongens Nytorv 6
DK-1050 Copenhagen K
Tel. (45) 33 36 71 00
Fax (45) 33 36 71 99
E-mail: eea@eea.eu.int
Internet: <http://www.eea.eu.int>

Contents

1. Background.....	4
2. Scope of the report	5
3. Quality control at the regional level	6
3.1. Quality control in sampling	6
3.2. Quality control in analysis	7
4. Quality control at the national level	10
4.1. Checks specific to individual data points	10
4.2. Generic checks	11
5. Quality control at the European level	12
5.1. Validation checks on Waterbase data	12
5.2. Upload to main working database	14
6. An index of quality	15
6.1. Background details to the components of the quality index.....	16
6.2. Application and uses of a data quality index	17
7. The way forward	20
References	21

1. Background

Eurowaternet is the process by which the EEA obtains the information on water resources (rivers, lakes, groundwater, transitional, coastal and marine waters, emissions and quantity). The process includes Eurowaternet reporting guidance documents, data exchange formats, annual EIONET priority dataflow, validation of data prior to compilation in working databases and finally publication of the harmonised European data sets in Waterbase on the web. It needs to answer the questions of its customers and to achieve its mission of 'supporting sustainable development and to help achieve significant and measurable improvement in Europe's environment through the provision of timely, targeted, relevant and reliable information to policy-making agents and to the public'.

Eurowaternet is based on a network of monitoring stations designed to provide the data required at European level relevant to EU legislation and policy development. It uses the monitoring networks that already exist in the EEA member countries and will be adapted to changes following the water framework directive requirements for monitoring.

Eurowaternet is a statistically designed stratified system, which, if implemented by countries in line with the guidance documents that are available, will give rise to

data that are comparable, with known statistical power and precision.

Generally, the data arising through Eurowaternet are used to compile the 'Indicator fact sheets', upon which the EEA reports and assessments are based.

Since the design, and pilot testing of Eurowaternet in 1996 and subsequent years, efforts have been addressed at implementation throughout the 31 current member countries of the EEA. The flow of data from each country through Eurowaternet (now an annual process performed electronically as part of the EIONET priority dataflows) and their storage in a relational database called Waterbase (which includes the previously separate Marinebase) has been regularly reported back to the member countries, the EEA Scientific Committee and the Board of Management. Water is one of the agreed priority data flows. Effort has been focused on the 'timely, targeted and relevant' aspects of the data. The EEA Scientific Committee and the Board of Management have drawn attention to the need to also address the reliability, or quality, of the data in Waterbase to ensure that assessments arising from them have a secure and defensible foundation.

This report addresses that stated need.

2. Scope of the report

Eurowaternet derives data from databases and information sources that already exist within the member countries. Generally, for rivers and lakes, the data are aggregated to annual averages (sometimes seasonally averaged data are required) and the aggregation is carried out by the member country itself from its 'disaggregated' national systems. In this context, therefore, the prime responsibility for quality of monitoring data lies with the member country. This responsibility is further reinforced through the process whereby the data in Waterbase, or in assessment reports using data derived from Waterbase, are passed to the national focal points for validation (quality checking) prior to publication. In the case of groundwater, the data are disaggregated but the onus of quality control and assurance still lies with the member country.

Because of differences between national systems in the ways in which data are stored and the substances are defined (e.g. nitrate may be described as NO_3 or N, phosphate as PO_4 or P), the ETC Water takes responsibility for the comparability of the data in Waterbase and carries out certain checks for correct values and conversions of different units to ensure comparability at the European level. The presumption has always been that the actual numerical data themselves are of good analytical quality.

This report provides some guidance to member countries on quality control procedures which need to be taken into account at the regional (sub-national) and national levels to provide the essential quality

base upon which all onward use of the data depends. The guidance is brief (but references are provided for additional, more detailed, information) and is focused on quality control of sampling and analysis. Tests on the integrity of the data carried out by the ETC Water before the data are uploaded to Waterbase are also described.

Until now, emphasis has been placed on physico-chemical determinands (in the water phase), and associated catchment characteristics and pressures of rivers, lakes and groundwater in the water phase. In future, attention will be turned to transitional, coastal and marine waters and biological and hydromorphological determinands as Waterbase expands. In addition, and especially in coastal and marine waters, data on contaminants in sediments and biota will need to be covered.

For transitional, coastal and marine waters, the standardised procedures on reporting performance in intercalibration exercises and in-house quality assurance procedures used by the marine conventions (especially OSPAR and Helcom) serve as a strong basis for quality.

Finally, an attempt to develop a semi-quantifiable index of quality of data held in Waterbase is described. The purposes of the Index are to allow users of Waterbase to make judgments about the between-country comparability of the data, and to stimulate countries to improve the quality procedures at regional and national level wherever necessary.

3. Quality control at the regional level

3.1. Quality control in sampling

Careful attention to the soundness of sampling and procedures for sample handling and preservation is essential if data of adequate accuracy are to be obtained. The cornerstone of effective sampling is the selection of suitable equipment and procedures, the provision of adequate training to sampling personnel and the design of the sampling programme so that natural variability is minimised. It is also necessary to ensure that appropriate control tests are applied to demonstrate the validity of the chosen procedures. A number of ISO guidelines on quality assurance have been produced and these are given in the list of references at the end of the report.

Control tests of sampling and sample handling have the same basic objectives as their counterparts in analysis, namely to ensure that any important deterioration of the accuracy of results, arising from these steps, is detected as rapidly as possible so that corrective action can be taken. In addition to general 'good practice' aspects of routine quality control in sampling (e.g. checks and preventative maintenance on sampling equipment), the specific control tests described below should be considered and put into practice wherever appropriate and practicable.

Ideally, sampling (and field measurements) should be carried out according to a written protocol (or standard operating procedure) by staff who have been trained. This forms the basis of question 1 in the index of quality (see Table 6.1).

(a) Routine tests on the effectiveness of the cleaning of sampling vessels and sample containers

Whilst field blanks (see below) give some check that such vessels and containers do not cause important contamination of samples, laboratory tests have the advantage that they can be routinely undertaken before sampling is performed; thus, if contamination problems are revealed, they can be rectified before sampling, thereby saving potentially wasted effort and resources.

(b) Field blanks to provide routine checks on contamination

Field blanks are samples of (typically) deionised or distilled water, which are taken into the field and treated, so far as possible, in exactly the same way as real samples. The exact details of the approach to be followed will, therefore, vary according to the particular system being controlled, but field blanks should generally be subjected to the same preparatory steps (such as filtration and centrifugation) as are applied to real samples, and should subsequently be handled, preserved and stored in the same way.

(c) Field check samples to provide routine checks on sample stability

In situations where, despite careful initial selection and testing of equipment and procedures, the stability of samples is in question, it can be useful to prepare check samples of known determinand concentration and treat them, so far as possible, in exactly the same way as real samples. Such a check sample may be prepared by dividing a typical sample into two and making a known addition to one portion. The recovery of the added determinand is a check that sample preservation, transport and storage are satisfactory and that loss of the determinand — by absorption or evaporation of volatile components, for example — is adequately controlled.

(d) Duplicate samples as an indication of sampling uncertainty

In conjunction with tests on analytical variability, the collection and analysis of duplicate samples can provide a check on the contribution of sample collection and handling to overall random error.

(e) A routine chart of field blanks may be a valuable way of monitoring control over sample contamination

Control samples of types (b) and (c) are similar to some of the analytical control samples described previously. Indeed, when analysed they will inevitably cover the sources

of analytical error controlled by those samples (as well as the potential sources of error in sample collection and handling that they are specifically intended to control). However, their use should not be regarded as a substitute for the use of the relevant analytical controls, because they can only be fully effective in controlling errors in sample collection and handling if the analytical process itself is under separate and effective control.

3.2. Quality control in analysis

The following sequence of activities is recommended as the basis of a technically sound approach to quality control in analysis of physico-chemical determinands in the water phase.

- (a) Definition of accuracy requirements that are consistent with the intended purpose of the analysis.
- (b) Selection of an analytical system capable of producing results of the required accuracy for the determinand in question. The analytical method should describe unambiguously, and in sufficient detail, the full analytical procedure.
- (c) Estimation of the within-laboratory total standard deviation of individual results for a range of sample types or matrices and concentrations representative of the samples and sample types of interest.
- (d) Estimation of spiking recovery achieved using the chosen analytical system for the sample matrix or matrices of interest.
- (e) Establishment of a documented, routine quality control system based on quality control charts, as a continuing check on analytical performance when the system is in routine use.
- (f) Participation in external interlaboratory quality control (proficiency testing) schemes involving the distribution of check samples.

3.2.1 Definition of accuracy targets

The following illustrates a logical general approach to be adopted for specifying the required accuracy of analytical results.

- (a) Two concentration levels should be defined for a given determinand. These are: (i) the lowest level of interest likely

to be encountered in the waters/sediments (**the minimum level of interest**); (ii) the concentration which represents the likely level at which most monitoring (e.g. for the assessment of trends or compliance with water quality standards) will be carried out (**the principal level of interest**). These two levels define the basis of the required accuracy. Experience suggests that it is usually appropriate to set a required limit of detection (C concentration units) which is at least one tenth of the principal level of interest and at least one third of the minimum level of interest. It is then necessary to select a tolerable percentage error (P %) that will apply to measurements made at concentrations near to the principal level of interest.

- (b) It is then assumed that the aims of the programme will be satisfied provided:
 - (i) that relatively few results are reported as 'less than' the minimum level — this will assist in load calculations and will ensure that real data are reported for the majority of sampling sites); (ii) **more importantly**, that the accuracy achieved at the principal level is not worse than $\pm 20\text{--}30\%$ of the principal level. The choice of a maximum tolerable percentage error, P %, will depend on the determinand — trace organic determinands might be allowed a greater tolerance to account for the fact that errors from many sources are possible.

These two levels define the aims of the programme; they indicate the performance needed from analytical systems. The relative importance of the minimum and principal levels of interest will depend on the purpose of the monitoring programme. At one extreme, a programme intended to evaluate the presence and relative distribution of a contaminant will tend to focus on the need to detect concentrations above the minimum level, and the achievement of the stipulated accuracy at the principal level may not be so important. At the other extreme, a programme of compliance monitoring for a determinand that is known to be present will require a clear demonstration that the required accuracy is achieved at the principal level of interest. The minimum level will be less relevant because most concentrations will be well in excess of it. Clearly, the whole philosophy will depend on an appropriate choice of the initial estimates of minimum and principal concentrations of interest.

More specific accuracy targets may then be defined as follows.

The total error of individual analytical results should not exceed C concentration units (e.g. $\mu\text{g l}^{-1}$) or P % of the result, whichever is the greater.

It is important when setting targets to allow, in the definition, for the existence of both random and systematic errors; the two types of errors have different effects on the use of and decision taken using analytical results.

The tolerable total error may be apportioned between error from random and systematic sources as follows.

‘The systematic error (bias) of individual analytical results should not exceed C/2 concentration units (e.g. $\mu\text{g l}^{-1}$) or P/2 % of the result, whichever is the greater.’

‘The random error of individual analytical results should not exceed C/2 concentration units (e.g. $\mu\text{g l}^{-1}$) or P/2 % of the result, whichever is the greater.’

A measure of the random error associated with analytical results is given by the standard deviation of results. The random error (95 % confidence limits) is equal to (approximately) twice the total standard deviation of analytical results. Thus if P = 20 %, it follows that the maximum tolerable total standard deviation, s is $0.25P = 5\%$.

3.2.2 Performance testing of analytical systems

It is necessary to establish the performance of an analytical system as a means of demonstrating adequate accuracy and as the basis for continuing routine quality control. The accepted approach for water laboratories is described in detail in ISO/TR 13530, Cheesman et al. (1989) and Timmerman et al. (1996). The minimum specifications for performance testing are summarised below. Although the following specifications are described in terms of determinations made on water samples, the same principles should be followed for analyses of other types of sample. For example, the same design of test is appropriate to determinations on solid sample, but the approach should be adapted to make use of standard reference materials. Newly established systems/methods should be tested before being used routinely. Existing methods should be reviewed at intervals of not greater than three years and

whenever procedures are modified. Performance testing should only be undertaken when the analytical system has been optimised.

3.2.3 Design of tests to evaluate performance

Tests should be carried out on a minimum of the following solutions. (These performance test solutions should be treated as samples, i.e. the calibration procedure would be that used routinely and would be implemented for each batch of analysis.) Determinations should be made in duplicate, in random order, in a sufficient number of analytical batches to provide at least 10 degrees of freedom for estimates of total standard deviation (11 batches of analysis guarantees 10 degrees of freedom, but fewer batches may be sufficient if between-batch variation is well controlled).

Box: Performance tests

- (a) A blank sample of water used to prepare the calibration standards (if no response is obtained for zero concentration samples, this should be spiked to a level where responses can be measured — say 0.05 of range) (note 1).
- (b) A standard solution of concentration at or near to the level of interest (e.g. EQS level) (note 2).
- (c) A standard solution of concentration reflecting the levels typically found in samples analysed routinely.
- (d) A natural water sample at 0.2 of range (spiked if necessary) (notes 4, 5, 6 and 7).
- (e) The same low concentration sample spiked to 0.8 of range (see above). This solution should be made up for each batch by spiking a portion of the sample analysed in (d) above (notes 4, 5, 6 and 7).

Notes:

- (1) Data from tests on sample (a) can be used to calculate a value for limit of detection (Analytical Methods Committee (1987)).
- (2) Standard test samples (b) and (c) should be prepared using material from a source separate from that of the calibration standards. These test samples should be made up freshly for each batch of analysis.
- (3) Bias can only be assessed with respect to a sample of known concentration (a reference material of suitably high specification or an independent check sample), but a preliminary check on calibration bias should be calculated as the difference between the mean and expected values for the standards (b) and (c) above.
- (4) ‘Range’ is used here to denote the range of likely interest, not the range over which the analytical technique is capable of operating. The effective range of interest is thus defined by the concentrations chosen for the test samples in (b) and (c) above.
- (5) For stable determinands (or those for which effective preservation techniques may be applied), a bulk sample should be collected,

preserved and used throughout the tests. For unstable determinands, two alternative approaches are recommended. For trace organic substances, a spiked sample may be prepared freshly for each batch of analysis, preferably using a water which contains a negligible concentration of the determinand. Alternatively, for nutrients and sanitary determinands a fresh sample of similar composition and concentration (0.2 of range) should be collected and used in each batch.

- (6) If the samples tested are of significantly different matrix from that analysed routinely, each matrix should be tested as indicated in (d) and (e) above to provide precision and recovery data. Results should be recorded in concentration terms as if the above solutions were routine samples.
- (7) For solid matrices samples, (a) to (c) would be the same, but standard reference materials of concentration low and high in the range of interest would be substituted for samples (d) and (e)

3.2.4 Routine quality control

Routine quality control in analysis is based on the use of control charts — see ISO (1991, 1993a and b), Analytical Methods Committee (1995), Thompson and Wood (1995) and Hunt and Wilson (1986). The laboratory should analyse a control sample at least once in each batch of analysis. The results of these control analyses are used to plot a control chart which is used to maintain the analytical system in a state of statistical control.

3.2.5 Control analyses

The control sample should be chosen such that it is subject to the same potential sources of error as samples analysed routinely. As a minimum requirement, the control sample should be a solution which contains a known concentration of determinand no greater than the level of interest. Where sample concentrations are greater than the level of interest, then additional control samples should be used to reflect sample concentrations. The type and frequency of use of control materials will depend on the analytical technique and the nature and likely sources of error which may affect results. Normally, between 5 and 20 % of all samples analysed should be control samples. All control samples should be subject to the full analytical procedure. The results for all control analyses should be recorded.

Where limit of detection is critical (e.g. for calculation of contaminant loads), duplicate blank determinations should be made in each routine batch of analyses. Limit of detection should then be re-estimated at 11-batch intervals from these measurements. Reporting limits should be based on the most recent estimate of limit of detection.

3.2.6 Control rules

It is essential that the laboratory has adequately documented procedures which define loss of statistical control and specific actions to be taken when an out-of-control condition arises. Records of breaches of the control rules should be maintained and, as a minimum, should include:

- (a) information to identify the control sample concerned and, via the batch of analysis, the identity of all associated test sample results;
- (b) details of the breach of control rules including a record of the control result and the control limits in force at the time;
- (c) action taken to investigate the cause of the out-of-control condition and any consequent conclusions and remedial measures;
- (d) action taken with respect to the associated test sample results.

The results of analyses obtained using a system not in statistical control should not be released, except under exceptional circumstances. Any such results should be identifiable for future examination and audit. The circumstances under which such results may be released should be documented clearly and include the specification that the cause of the out-of-control condition should first be identified and shown not to affect results for the analysis of samples. The control chart should be reviewed periodically and the control limits updated if necessary. The results of all quality control analyses should be taken into account in calculations of performance and in updating charts, apart from out-of-control values for which the cause has been identified and remedied.

3.2.7 Interlaboratory quality control exercises — proficiency tests

Unless it is agreed otherwise, the laboratory should adhere to the test protocol for an interlaboratory exercise. Samples provided in proficiency testing schemes should be treated as far as is possible in the same way as routine samples with respect to storage, registration, analysis and reporting. Routine AQC procedures should be applied. In particular, any replication of analysis carried out as part of interlaboratory test should be 'blind'. Individual replicates need to be submitted for analysis independently and without reference to one another. No more than the specified number of determinations should be made.

4. Quality control at the national level

Usually, this will be primarily concerned with data validation and screening processes aimed at producing a common, homogenous (i.e. free of regional variations) national set of data.

Data 'validation' or 'screening' procedures for individual data points fall into different categories. The list below indicates the main types.

4.1. Checks specific to individual data points

Data screening procedures applied to individual data points fall into two main types, as listed below.

(a) Logical

This type of screening procedure checks the data point against logical tests — the reported value is evaluated against a specific rule that is applied to all data of a given type. Examples include:

- pH value — 'Is the reported value between 0.0 and 14?' (NB: values outside this range can occur in extreme circumstances or non-aqueous solvents, but the range given applies to natural waters);
- BOD — 'Is the reported value less than that reported for chemical oxygen demand (COD)?';
- dissolved metal — 'Is the reported value more than 1.2 times that reported for total metal?' (NB: the factor of 1.2, or similar is included to allow for the effects of random analytical error).

(b) Statistical

This type of screening procedure tests the current data point against the corresponding data set reported previously. Values that appear to arise from a different statistical population are queried. This approach is particularly valuable for screening for outliers.

Occasions can arise where a data set contains a suspected 'outlier' — a data value which has arisen from some statistical population that is more extreme than the population from which the bulk of the values have

arisen. The problem of outliers arises in all areas of environmental monitoring.

An outlier can be identified **only in relation to an assumed probability distribution**. Thus, unless there is very firm evidence to support the assumption of, say, a logNormal population, there is no way of telling whether an apparent outlier is (i) the result of a sampling, analytical or recording error, or (ii) valuable evidence that the logNormal assumption is invalid.

In investigating a suspected outlier, the first step should be to follow the audit trail back and check that the outlier is not merely due to a recording error. Practical range checks based on simple science should also be made at this stage: for example, a river water pH of 17.2 or a groundwater temperature of 159 °C would clearly be wrong and could reasonably be scrapped. Next, if a portion of the sample has been retained, this could be reanalysed to check whether analytical error had caused the problem. Finally, if the exact location of the sample could be identified, resampling could be an option.

What if, despite all these checks, nothing untoward is uncovered? There is a temptation to assume that the outlier is a rogue result anyway and that it may reasonably be discarded. When doing a **retrospective analysis** of a historical body of data, this attitude **can** often be justified — especially if the outlier occurred in the distant past (since which time sampling and analytical procedures may well have improved). In particular, it is appropriate to discard unattributed outliers if the aim is to obtain a 'best-case' measure of variability under stable conditions of quality.

For **ongoing** monitoring, however, the routine discarding of inconvenient or awkward outliers **is not an acceptable option**. In the context of the present guidance document, there is a very real possibility of hot spots in the data — and this is precisely the phenomenon that can be guaranteed to give rise to suspected outliers! To suppress such evidence without very sound justification could lead to highly misleading conclusions being made.

4.1.1 Practical examples

The general aim is to clean up a historical set of data prior to estimating baseline statistics. If outliers are not identified and removed, where appropriate, estimates of standard deviation can be grossly inflated, and this will lead to inappropriately insensitive assessment values.

Specific problems that can be flagged by such analysis include:

- sites or determinands where concentrations are being recorded or coded in a mixture of units — e.g. mg/l and mg/l;
- cases where contamination is introduced in samples taken by an inexperienced sampler;
- a rogue set of results from the laboratory, due to a sample bottle wrongly labelled or an incorrect standard used.

4.1.2 Limitations

The recommended method makes the assumption that the data are normally distributed (after logging where appropriate). For outlier detection the assumption is particularly critical, as the method is concerned with the extreme tails of the distribution — which is precisely where

the assumption is most likely to break down. For this reason, the outlier test should be regarded as providing no more than a rough screen of the data, with an element of judgment applied in marginal cases. Nevertheless, experience shows that outlier tests are extremely useful for flagging up gross outliers (such as those in error by a factor of 1 000) and, in general, the routine use of such tests as a preliminary to the main statistical analysis is highly recommended.

4.2. Generic checks

Generic assessments of data quality are those that apply to all results reported from a particular source. These are less widely applied at present, but are being developed to assess the fitness for purpose of data produced for environmental monitoring programmes.

Data quality assessments providing more of an overview of fitness for purpose focus on the procedures implemented by laboratories in respect of quality assurance and quality control procedures. Such procedures are rated in relation to the extent to which they reflect best practice and ensure that the standard of accuracy achieved for monitoring data can be demonstrated.

5. Quality control at the European level

As previously stated, the prime responsibility for data quality lies at the regional/national levels, as the Eurowaternet criteria require member countries to provide only nationally validated (and usually aggregated) data.

Efforts at the level of the European Topic Centre on Water are therefore directed towards screening for outliers and transpositional errors which may have got through national screening procedures or may have been introduced during the transfer from national to European level.

5.1. Validation checks on Waterbase data

5.1.1 Have we got the right data?

The ETC Water often receives a mixture of river/lake/groundwater/quantity station characteristic/quality/pressure data in aggregated/disaggregated formats. These data sets need to be distributed to the appropriate ETC Water partner, i.e. rivers data to WRc, lakes to NERI and groundwater to AWW.

For the rivers update, WRc should receive at least three files of data: stations (physical characteristics), pressures and chemicals (aggregated). What is actually received varies from country to country. In many cases, only the quality data are received and the station IDs have to be extracted from the quality update and the physical characteristic and pressure data re-requested. Pressure data are very patchy.

Some countries have only provided disaggregated chemical quality data. In these cases, the data are aggregated before uploading to the database.

For the lakes update, the process is essentially the same as for rivers except so few pressures data have been provided that this has been temporarily excluded from the database.

For the groundwater update, the data comprise: groundwater (GW) bodies (a list of bodies including some key characteristics), a characterisation of each GW-body and quality data (disaggregated or aggregated data). Characterisation data can be provided either

online by accessing the working database and delivering the data via an Internet form or by using an Excel template. Quality data can be provided in aggregated form by using an Excel template or in disaggregated form as a text file according to the interface description.

What is received varies from country to country. If data provision comprises groundwater body characterisation only and if this is provided online then no data file is received. In the case of the provision of filled-in Excel templates, the maximum number of files is one file for the list of GW bodies plus two files (characterisation and aggregated data) for each GW body where information is provided.

5.1.2 Are the data in the correct format?

For rivers, the 'correct format' is a return of completed Excel templates supplied to each country with the last data request. In the majority of cases, data are not provided in this format. Data arrive in various formats including Access databases (especially if very large numbers of records are supplied). Most frequently, Excel spreadsheets are provided, but in the format of the country's choosing. Other countries send text files (.csv and .txt), data embodied in e-mails, and Word tables. The data are reformatted to fit three temporary tables in the working database which are called by various scripts to upload the data into the main chemicals, stations and pressures tables. These three temporary tables, temp_chemicals, temp_stations and temp_pressures are in the same structure as their respective main tables. In order to reformat the data, each country's files are imported into a pre-upload working database and a series of checks are systematically made.

For lakes, the situation is very similar to rivers but the preferred format for data deliveries is tab-separated, although Excel spreadsheets can also be handled easily.

Groundwater data which are delivered in Excel templates are extracted by VBA-macros and transformed according to the interface description before uploading the data into the working database. Therefore it has to be

checked whether the Excel forms are still in the original format (although the templates are locked, a lot of manipulations take place, year by year). For disaggregated data which are provided as text files, it has to be checked whether the data are delivered according to the interface description or not.

5.1.3 Pre-upload checks

For rivers, before the raw data are imported into the pre-upload working database, the supplied files are opened in Excel or WordPad (depending on size) and any non-numeric characters such as asterisks and hyphens are removed from numeric fields. Any commas used to represent decimal points are replaced with a full stop. The fields are renamed to the standard Waterbase fieldnames and any missing fields are inserted, such as year or determinand. A copy of the amended Excel or text file is saved in each country's working directory. The raw data (as delivered) are also stored under each country and are never overwritten.

Once these initial checks have been made, the files are imported to the pre-upload working database.

The determinands and units are then checked. They are updated to the standard terminologies and any appropriate correction factor is applied to the value fields (e.g. for converting NO₃ to N and PO₄ to P). Any parameters supplied that are not listed in the standard determinands table (such as suspended solids, depth, heavy metals) are removed and held in a separate table.

The next series of checks are used to determine whether any data are missing. A series of queries are made between the stations supplied in the update (both from the stations and quality update tables) and the existing main stations table, and also between the stations and chemical quality tables provided within the update. Most countries seem to supply either a new station with no quality data or quality data for a station with no characteristic detail.

There are several scenarios.

1. Missing stations

Stations are listed in the main stations table but there are no physical characteristic data in the update. The main stations table is checked against the stations and the quality update, depending on what has been

supplied. Stations and quality update are also checked between themselves.

2. New stations

Stations provided in the update (stations and/or chemicals update) are not in the existing main stations table. There is a need to request physical characteristic and pressure data for the new stations if they are not provided in the update.

3. No change

The stations in the update (chemicals and/or stations) match those in the existing main stations table. Changes to details are checked on upload to the main tables.

A list of queries is produced for each country following the stations checks — these are referred back to the NFP/NRCs for clarification.

For lakes, the pre-upload checks are similar to those for rivers, apart from these details:

- the raw data are converted manually (using Excel, WordPad or other tools) to a format readable to SAS; not necessarily same format as the working database;
- SAS programmes/scripts read data and restructure them to fit into the working database structure (SAS is very powerful for such things); they also convert data to standard units and standard determinands;
- temporary data tables are generated in each country's folder;
- temporary tables are checked as described for rivers.

For groundwater, before the raw data are imported into the pre-upload working database, the supplied files are opened in Excel or WordPad (depending on size) and any non-numeric characters such as asterisks and hyphens are removed from numeric fields. Any commas used to represent decimal points are replaced with a full stop. The fields are renamed to the standard Waterbase fieldnames and any missing fields are inserted, such as year or determinand. A copy of the amended Excel or text file is saved in each country's working directory. The raw data (as delivered) are also stored under each country and are never overwritten.

Once these initial checks have been made, the files are imported to the pre-upload working database.

5.2. Upload to main working database

Rivers data are appended to the three temporary tables mentioned earlier and exported to a folder on the main ETC drive, from where they are imported to the main working database for upload to the main tables.

On upload, a comparison routine is run whereby each field being updated (i.e. already existing in one of the main tables) is checked and any changes are written to an archive table first, to retain an audit trail of amendments. Fields are not amended to null if a value already exists. Any new records are appended to the main tables.

Lakes data make use of SAS scripts to upload data from the temporary country tables to the common working database tables. If the countries have made a complete re-delivery, old data are replaced, otherwise appended.

For groundwater, the data files which were prepared according to the interface description are uploaded to the working database. Several data checks are part of the upload procedure.

The format of data (strings, integer, date, etc.) as well as the existence of GW bodies corresponding to the characterisation and quality data is checked in a first step of the upload procedure. It is mandatory to draw up groundwater bodies first before uploading general characterisation data or quality data.

In the case of failures, the upload procedure stops before importing the data into the database and lists all the failures in an import report.

In case of data upload, existing data are overwritten. At each upload process, the uploaded data are marked by a transaction code. This transaction code is very helpful in reversing whole upload cycles.

After finishing the upload process successfully, an import report is generated by the working database, giving the transaction number, the number of inserted, updated and erased values and the number of transactions.

Outlier checks are then performed by running query tools which are part of the working database.

6. An index of quality

To respond to the steer from the EEA Scientific Committee and the Management Board, the ETC Water is proposing a system whereby the data in Waterbase can be assigned a 'quality index' which will quantify the quality stages that the data have passed through from sampling, analysis and reporting.

It is proposed to ask each national focal point (or more realistically the appropriate national reference centre) a series of questions which will allow an objective assessment to be made about the comparative quality of the data provided through Eurowaternet. This will be done for each data set (i.e. rivers, lakes and groundwater). It has been presumed that regional quality procedures will be harmonised to a national standard, so the questions need answering only at the national level. The questions will need to be asked at each annual update to take account of any national changes (improvements)

The relative scores assigned to each category are intended to reflect the importance of the subject to ensuring that monitoring data are likely to be fit for their intended purpose.

The highest score is assigned to the use of routine quality control techniques (question 6) since this is fundamental to the maintenance and demonstration of measurement quality. The definition of accuracy targets for a monitoring programme (question 4) and the evidence of quality provided by independent interlaboratory tests (question 7) are ranked next in order of importance. Finally, the organisational aspects of sampling, accreditation, external audits, performance testing and data screening are included in the assessment scheme.

All the questions in the index of quality refer to participation in quality assurance (QA) procedures rather than performance in QA procedures. It has been presumed that poor performance in QA procedures at regional or national level will have been detected and corrective action taken before data are accepted into the national databases. See Section 6.2 for details on how the quality index may be applied.

The highest quality index is 12 and the lowest is 0.

Questions to ask about national or regional monitoring programmes

Table 6.1

	Sampling	Score, if yes
1	Is sampling (and are any field measurements) carried out to a documented protocol by staff who have undergone specific training? (See Section 3.1.)	1
	Analysis	
2	Are the analytical laboratories accredited by a national accreditation body — to ISO 9000 or EN45000 series standards? (See Section 6.1.)	1
3	Are the laboratories subject to external audit? (See Section 6.1.)	1
4	Have numerical accuracy requirements been defined for all relevant determinands? (See Section 3.2.)	2
5	Do laboratories have performance test data for their own analytical systems — indicating the precision of analysis, spiking recovery and limits of detection? (See Section 3.2.)	1
6	Can the laboratories produce routine quality control charts for all relevant determinands? (See Section 3.2.)	3
7	Is the monitoring programme linked to a series of routine and regular interlaboratory tests — for all relevant determinands either on a national or international basis? (See Section 3.2.)	2
	Data screening	
8	Are the monitoring data automatically (i.e. using specific software) screened for statistical outliers or checked for unusual results before being stored on a national or regional database? (See Section 4.2.)	1

6.1. Background details to the components of the quality index

6.1.1 Sampling

Use of adequately trained sampling staff who are working to a defined sampling protocol is essential if sampling is to be carried out in a consistent and valid manner.

6.1.2 Accreditation

The fact that a laboratory is accredited provides important evidence that its work is carried out in a controlled and well-organised manner. Several international standards have been produced which define the concepts of and approaches to quality assurance. The most general of these is ISO 9000 — Quality systems (European Standard 29000). Other more specific standards give details of how to implement the principles of quality assurance in different situations. The standard of principal concern in chemical analysis is ISO Guide 25 — General requirements for the technical competence of testing laboratories. This guide has achieved wide acceptance and has become the generic standard relating to laboratory accreditation (see below). The guidance given in ISO Guide 25 is expanded upon in a series of European Standards: EN45001 — General criteria for the operation of testing laboratories; EN45002 — General criteria for the assessment of testing laboratories; and EN45003 — General criteria for laboratory accreditation bodies. These standards define the important aspects of a quality system which would be required in order to ensure that analytical results are fit for their intended purpose. These criteria also act as the basis on which to identify a competent laboratory. This idea has been developed in many countries into the concept of ‘accreditation’. Accreditation for a testing laboratory is the formal recognition (by a nationally or internationally recognised authority) of the competence of a laboratory to carry out certain tests. The standard of competence is defined partly by having a clear specification of the laboratory’s organisational and record-keeping requirements (quality system) and partly by ensuring that the standard of accuracy achieved is demonstrably adequate for the intended application. To be certain that fitness for purpose is achieved (rather than merely assumed), there is a requirement for accredited laboratories to take steps to determine their customers’ needs.

6.1.3 External audit

External audit is a feature of accreditation. However, laboratories that may not be accredited can take steps to allow their operation to be examined by key customers, regulatory bodies or other interested external organisations. The fact that they are subject to external audit means that they will need to have documented procedures and is an indication of attention to quality.

6.1.4 Clearly defined accuracy targets

Well-designed monitoring programmes will specify a required level of accuracy in a clear and unambiguous manner, usually in terms of a numerical maximum tolerable error. This should be defined in such a way as to apply across the whole concentration range of interest. Clear accuracy targets are important in terms of providing a numerical expression of fitness for purpose and hence a criterion for analytical performance. This in turn assists in the selection of suitable analytical techniques and the design of within- and between-laboratory quality control activities.

6.1.5 Laboratory performance test data

Analytical methods do not possess an inherent performance of their own, only the capability to perform up to limits defined by the technique on which they are based. In practical use, no analytical method will achieve its theoretical performance. This is because of a wide range of factors including: sub-optimal design of practical methods, errors or inconsistencies in the application of the analytical method, imperfections in instrumental components, human errors by operators, environmental variables, etc. These factors mean that any measurement will be subject to error and that errors will vary from one measurement system to another, even if the same method is used. If the errors associated with analytical results were always very small (and it were certain that sample handling and the reporting of data were always totally reliable), the correctness of data interpretation and any consequent decisions would rarely be in doubt. However, many experimental studies have shown that analytical results are often subject to serious errors, particularly at the low concentrations encountered in water analysis. Hence, for all applications of analytical data, it is important to assess the performance of analytical systems in relation to the level of accuracy required.

6.1.6 Routine quality control charts

The maintenance of routine quality control charts is the mainstay of an effective within-laboratory quality control system. Within laboratory quality control has the aim of demonstrating that each batch of analysis is carried out in a controlled manner that is consistent with the project aims. To be fully effective, interlaboratory tests should be supported by a programme of within-laboratory quality control. The organisation of a programme of interlaboratory quality control should be seen as the final component of the overall approach to the achievement of adequate data quality. The essential features are that laboratories should carry out some form of control analysis in each analytical batch, that the control analysis is used as a means of assessing the validity of the batch of analysis and that the control data should be plotted on a statistically based chart.

6.1.7 Interlaboratory tests

A programme of interlaboratory tests has the purpose of providing an independent demonstration that laboratories involved in the monitoring programme can undertake the required analysis and meet the accuracy targets. Experience indicates that it is worth emphasising the point that participation in interlaboratory tests is insufficient; laboratories should be required to take part and to perform adequately. Some means of monitoring performance and of assessing levels of performance should be part of any well-organised monitoring programme.

6.1.8 Data screening

Automated data screening for statistical outliers or results that depart markedly from historical data is a valuable means by which erroneous results can be eliminated from the monitoring programme database.

6.2. Application and uses of a data quality index

The text and table below are intended as an illustration of how a quality index for monitoring data might be applied to environmental data. The quality index is a score that refers to aspects of quality assurance that are associated with the generation of the data — the higher the index, up to a maximum of 12, the greater the confidence that can be placed in the data. It should be the responsibility of the user of data to examine the quality index, including individual components to check that criteria are consistent with the requirements of the data application in hand.

A high quality index score indicates that data should satisfy the needs of most applications where it is required that:

- (a) data have been obtained in a consistent manner using widely accepted measurement techniques;
- (b) current norms of analytical performance have been achieved; and, most importantly,
- (c) that all data are accompanied by information that can be used to demonstrate (a) and (b) above.

Lower scores do not necessarily mean that data are of low quality (see below). A low score is an indication that data may be less reliable, usually because a fully adequate demonstration of fitness for purpose is not available. Monitoring data produced in a research context or as part of 'one-off' surveys often do not meet all the quality index criteria — and hence do not qualify for a maximum or near maximum score.

The quality index may be used in selecting data for use in a particular application. The data user might need to consider the quality index threshold below which the data should not be used or should be accorded provisional status. The level of the threshold will, of course, vary for different applications. For example, one of the most demanding uses of data is the comparison of environmental quality between different countries or regions (when data have usually been submitted by various different laboratories). In such cases, where international comparability of data is the issue, it would be prudent to accept only data of a high quality index, perhaps greater than, say, 8. An example of a less demanding application might be the examination of a trend in environmental quality in a single region (where data have been submitted over time by the same laboratory). Here, internal consistency is the main issue and a lower score (one that focuses on the routine aspects of quality control such as charts) may be admissible. It is doubtful that determinations having quality index scores of 5 or lower can be regarded as being of demonstrable quality. However, there are some situations where no other data are available. In these cases, the data user must consider the usefulness of the data and the reliability of any conclusions drawn as a consequence. It is worth noting at this point that the data quality index can be an important (and useful) justification in data analysis for choosing not to accept all analytical data at face value and as of equal weight.

Table 6.2 is intended to provide examples of the quality index that would be assigned to a selection of environmental determinations undertaken as part of UK monitoring programmes. UK data have been used because the information concerning extent of quality-related activity was available.

Examples 1 and 2 (nutrients and metals in the national marine monitoring programme) are from an established, centrally coordinated, national monitoring programme and might be considered to refer to data of the highest quality and most demonstrable reliability. (Note, the maximum score of 3 for the use of control charts has been reduced to 2 because not all contributors to the programme have charts.) Example 3 (PAHs in sediment) is from the same programme but relates to a type of determination that, in a number of laboratories, is under development — hence a lower score. Example 4 (BOD in effluents) refers to a determination that, in the UK, is not centrally coordinated, but which because of its environmental and commercial

importance is closely controlled, documented and checked. Example 5 (reactive aluminium) is an illustration of a determination that until recently has been carried out in a research context, but now is under consideration as a routine determinand. This status is reflected in the facts that the tests carried out currently are subject to routine control and that performance is relatively well-characterised. However, a routine interlaboratory check on the determination is yet to be established. Finally, example 6 (oestrogens in water) illustrates the score that might be assigned to a research determinand for which the only source of data might be one-off research reports in the scientific literature.

Table 6.3 addresses data arising from the UK's general river quality assessment (GQA) monitoring programme which is the main data source for Waterbase rivers.

Table 6.4 shows the results of applying the data quality index in seven EEA member countries.

Table 6.2 Data quality index — Examples from UK monitoring programmes

	Example No		1	2	3	4	5	6
Questions	Examples of determinations undertaken for environmental monitoring in the UK	Possible score, if yes	UK NMMP Nutrients in seawater	UK NMMP Metals in sediments	UK NMMP PAH in sediment	BOD in sewage effluents	Reactive Al in UK surface waters	Oestrogens in UK rivers
Sampling 1	Is sampling (and are any field measurements) carried out to a documented protocol by staff who have undergone specific training?	1	1	1	1	1	0	0
Analysis 2	Are the analytical laboratories accredited by a national accreditation body — to ISO 9000 or EN45000 series standards?	1	1	1	1	1	0	0
3	Are the laboratories subject to external audit?	1	1	1	1	1	0	0
4	Have numerical accuracy requirements been defined for all relevant determinands?	2	2	2	2	0	2	0
5	Do laboratories have performance test data for their own analytical systems — indicating the precision of analysis, spiking recovery and limits of detection?	1	1	0	0	1	1	1
6	Can the laboratories produce routine quality control charts for all relevant determinands?	3	3	2	0	3	3	0
7	Is the monitoring programme linked to a series of routine and regular interlaboratory tests — for all relevant determinands either on a national or international basis?	2	2	2	2	2	0	0
Data screening 8	Are the monitoring data automatically (i.e. using specific software) screened for statistical outliers or checked for unusual results before being stored on a national or regional database?	1	0	0	0	1	0	0
	Quality index score	12	11	9	7	10	6	1

NB: NMMP = national marine monitoring plan.

Data quality index — Example based on the England and Wales Environment Agency's general quality assessment (GQA) monitoring programme

Table 6.3

Questions	Examples of determinations undertaken for environmental monitoring in the UK	Possible score, if yes	EA GQA rivers data	Comments
Sampling 1	Is sampling (and are any field measurements) carried out to a documented protocol by staff who have undergone specific training?	1	1	The EA has the <i>National sampling procedures manual</i> , which documents a series of procedures and work instructions on how samples must be taken. There is a system in place in areas where designated staff are trained as trainers to ensure all sampling staff use the correct procedures. Proper training is also part of the staff competency requirements.
Analysis 2	Are the analytical laboratories accredited by a national accreditation body — to ISO 9000 or EN45000 series standards?	1	1	OK
3	Are the laboratories subject to external audit?	1	1	Yes, by UKAS
4	Have numerical accuracy requirements been defined for all relevant determinands?	2	2	The National Laboratory Service has specific target requirements for accuracy and precision for all determinands they analyse.
5	Do laboratories have performance test data for their own analytical systems — indicating the precision of analysis, spiking recovery and limits of detection?	1	1	OK
6	Can the laboratories produce routine quality control charts for all relevant determinands?	3	3	OK
7	Is the monitoring programme linked to a series of routine and regular interlaboratory tests — for all relevant determinands either on a national or international basis?	2	2	Yes, via AQUACHECK
Data screening 8	Are the monitoring data automatically (i.e. using specific software) screened for statistical outliers or checked for unusual results before being stored on a national or regional database?	1	1	Data are screened but the final data set still has errors of duplication and outliers. Better screening process to be used.
	Quality index score	12	12	

Examples of testing the data quality index in several EEA member countries

Table 6.4

		Score	DK	ES	FR	HU	NO	PL	UK (E & W)
Questions	Question/waterbody type (R = rivers; L = lakes; GW = groundwater; ALL = inland, transitional, coastal and marine waters)		ALL	R, L, GW	R, L, GW	R, L, GW	ALL	R, L	R
1	Is sampling (and are any field measurements) carried out to a documented protocol by staff who have undergone specific training?	1	0	1	1	1	1	1	1
2	Are the analytical laboratories accredited by a national accreditation body — to ISO 9000 or EN45000 series standards?	1	1	1	1	1	1	1	1
3	Are the laboratories subject to external audit?	1	1	1	1	1	1	1	1
4	Have numerical accuracy requirements been defined for all relevant determinands?	2	0	2	2	2	2	0	2
5	Do laboratories have performance test data for their own analytical systems — indicating the precision of analysis, spiking recovery and limits of detection?	1	1	1	1	1	1	1	1
6	Can the laboratories produce routine quality control charts for all relevant determinands?	3	3	3	3	3	3	3	3
7	Is the monitoring programme linked to a series of routine and regular interlaboratory tests — for all relevant determinands either on a national or international basis?	2	2	2	2	2	2	0	2
8	Are the monitoring data automatically (i.e. using specific software) screened for statistical outliers or checked for unusual results before being stored on a national or regional database?	1	1	0	0	1	1	1	1
	Quality index score	12	9	11	11	12	12	8	12

7. The way forward

The approach taken and the examples given above are based on rivers and are from the United Kingdom and six other EEA member countries (partners of the ETC Water). The approach appears to give a robust and useful indicator of the quality of the national data held in Waterbase and allows comparisons to be made between countries.

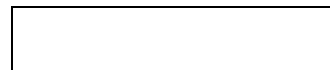
The data quality index was discussed at the EEA/ETC-EIONET workshop in Budapest in May 2003 and the overall conclusion was that the index is a driver of quality at regional, national and European level and that its further development should continue.

The next steps are to:

- apply the quality index to lakes and groundwater data from all EEA member countries;
- assess the results and check if the weightings to the questions are sufficiently sensitive to demonstrate real quality differences — for example, some countries have suggested that higher weighting be given to sampling (question 1) and data screening (question 2);
- consider the feasibility for transitional, coastal and marine waters (where the position is more complicated because of the involvement of the various marine conventions and their quality screening programmes);
- adapt the quality index to cover hydromorphological and biological data (this is dependent on data flows of these determinands being made available by the member countries);
- Report back to the EIONET group (especially the national reference centres) on progress made during 2003 to seek further advice and support.

References

- Analytical Methods Committee (1995), Statistical Sub-group of the Royal Society of Chemistry, 'Internal quality control of analytical data', *Analyst*, 120, pp. 29–35.
- Analytical Methods Committee (1987), Statistical Sub-group of the Royal Society of Chemistry, 'Recommendations for the definition, estimation and use of the detection limit', *Analyst*, 112, pp. 199–204.
- Cheeseman, R. V., Gardner, M. J. and Wilson, A. L. (1989), *A manual on analytical quality control for the water industry*, WRc report NS30, ISBN 0902156853.
- Hunt, D. T. E. and Wilson, A. L. (1986), *The chemical analysis of water, general principles and techniques*, Royal Society of Chemistry, London, 1986.
- ISO/TR 13530 (1998), Water quality — Guide to analytical quality control for water analysis.
- International Organisation for Standardisation (1991), ISO 8258 — Shewhart control charts.
- International Organisation for Standardisation (1993a), ISO 7870 — Control charts, general guidance and introduction.
- International Organisation for Standardisation (1993b), ISO 7863 — Control charts for arithmetic mean with warning limits.
- Thompson, M. and Wood, R. (1995), 'Guidelines on internal quality control in analytical chemistry laboratories', *Pure Appl. Chem.*, 67, pp. 649–666.
- Timmerman, J., Gardner, M. J. and Ravenscroft, J. E. (1996), UNECE Task Force on Monitoring and Assessment, Volume 4 — Quality assurance, ISBN 9036945860.



European Environment Agency

Eurowaternet: towards an index of quality of the national data in Waterbase

2003 – 21 pp. – 21 x 29.7 cm

ISBN 92-9167-578-4