

## DATA REVIEWS AND AUDITING

Data reviews are reports presenting the results of a data review process. Each report, and hence its reviewing process, is subsequently audited for conformance to standard and for accuracy, using a number of checking processes and having regard to the known and reasonably expectable behaviour of the parameters measured.

When certified as “passed audit”, the data are considered to be fully updated and, subject to the information contained in the filed comments, in a “confidently usable” state. The only qualification to this is that some interpreted data, such as flow ratings, may subsequently be updated if additional data becomes available.

### CONCEPTS AND GENERAL PROCEDURES

1. The purpose of the data review and its audit is to:
  - facilitate the process of checking that the data meet the standards required by the quality system
  - certify that the data are “confidently usable” by clients and other potential users, (subject to the information contained in the filed comments), and have no apparent and unexplained anomalies
2. The purpose of the data review *document* is to demonstrate to the auditor that the comprehensive checking process has been thoroughly and competently carried out for the above purposes, and to record this.
3. The review document is for internal use only. However at the Branch office’s discretion it may be made available to clients, and should then contain additional data, header pages, explanations, etc., and possibly additional data.
4. In carrying out the checking and auditing processes we are looking for both anticipated and unanticipated errors. Therefore the questions posed in the following sections should not be focused on exclusively.
5. The review document should be retained as a quality record, and may be in the form of a bound report, or in a file compiled for the purpose.
6. The checklist used to audit the review document by the auditor (either Form ADR or RDR) has the site and review details on it as well as a Conformance to Standards table. It is bound in as an integral part of the review document. A copy of the first page of this form shall be provided by the auditor, following the audit being passed, to the QA project leader.
7. Auditors’ findings will be noted on the form, as “conditions”, “recommendations” or “notes”. Similarly to other internal audits, “conditions” must be done and reported back to the auditor for the review to pass. “Recommendations” will be those that are seen as desirable, but are either minor (such as mis-spelt comments), outside the scope of the review or non-essential enhancements. (It would be expected that recommendations would eventually be carried out,

especially if noted in several audits). “Notes” are used for the assistance of the branch or the auditor.

8. Checking processes used in the review will include checks against other stations using suitable statistical methods which may be those employed by users of the data or may be specially written for the purpose. It is expected that these methods will develop over time. The scope of a data review is to cover all data including such items as comments that have a bearing on the quality and interpretation of the data for the period and immediately prior to it.
9. Each five years, one-fifth or more of stations shall have the data from the last 5 years or more reviewed and reported upon. Following the passing of this review by the auditor, the previous annual data reviews may be replaced by this one and destroyed.
10. The contents of the review documents are an attempt at answering the question “do the data meet the standards as given in the quality manual?” Given that every data point cannot be independently checked, this question cannot be answered with absolute certainty; however statistical analysis allows us to estimate the data uncertainty. And as our analysis tools improve over time, the review procedures will continue to evolve.

### **COMPILING THE REVIEW**

Following the quarterly dispatch of data to the Water Resources Archive by early February each year, the annual task of compiling the reviews can begin.

1. Firstly, a cover page should be compiled, comprising as a minimum the station name, the review period, branch, and “Preface” text on this or a following page.
2. The Contents page follows, comprising as a minimum the standard requirements for the type of station, as given in Appendix F. Header pages between each section explaining the data presentation and the significance of the parameters are optional, but should be included if the report is to be provided to clients.
3. The review form (SDR or RDR) follows. The preliminary details and the Conformance to Standards table need to be filled out by the reviewer. The values for the last 5 year period should be filled in from the data outputs and the team’s records if necessary. Fill in all years for both the 5-yearly review and the annual reviews. These data are required to indicate trends and will also be used for overall quality management. They can be compiled from:
  - gauging % are available from the QRAT outputs (last 5 years)
  - missing record % can be derived from Process GAP or script files
  - % of time the stage resolution is outside  $\pm 3$  mm refers to any period of record from a lower resolution instrument (the proportion of time that stage is estimated to be measured to a lower accuracy because of surge, etc., at high flows is another statistic, not considered at present)
  - the % of time that time resolution is outside half of the recording interval refers to any period of the record from an instrument not capable of this.
4. The data summary pages from Tideda, QRAT, etc., are compiled in the order of the contents page, with or without header pages as applicable.

Amplification of the standard requirements for flow stations is given below, with questions for the auditor in italics:

1. Cover, contents page and review form with its Conformance to Standards listing. The latter should be compiled from the data summaries rather than the HA quarterly reports. (Note that non-achievement of these standards will not prevent the review passing audit.)
2. Comments. A listing of station comments (stage, gaugings, ratings, rainfall, as applicable) for the full period of record are required to answer the questions:
  - *do initial comments provide the user with sufficient information to understand the features, accuracy standards, peculiarities and limitations of the data, etc?*
  - *are they in our standard format (i.e. adequate)?*
  - *are all anomalies in the data such as missing record, below-standard data, unusual features, gaugings outside  $\pm 8\%$ , etc., commented adequately?*
3. Listing of gaps from Process GAP. This should be compared with the listing of comments to answer:
  - *are all the gaps commented?*
  - *do all the gaps appear to be justified?*
  - *do the gaps for the period reconcile with the Conformance to Standards MR value?*
4. Listing of ratings. This is to cover the period under review plus the rating previous to the first applicable. Questions are:
  - *are all gaugings filed correctly?*
  - *do they cover the full range of stage during the period? (check PEX; also check the stage plots, noting the rating changes from the “bed plots”)*
  - *are adequate co-ordinates filed to describe the shape (especially at bends where bank overflow occurs, etc.)? (check against rating plots)*
5. Rating statistics for the last five years of record, plus the statistics for the full period
  - *have any rating changes been missed?*
  - *do at least 95 % of the gaugings lie within  $\pm 8\%$  of the rating?*
  - *do any ratings have statistics showing a significant bias and/or a non-random distribution of gaugings such that the rating needs to be corrected?*
  - *do any flow segments have significant bias and/or non-random distributions such that curves may need to be re-drawn or rating changes may have been missed?*
6. Extremes - PEXTREMEs of stage and flow for the full period (using the “Annual” option) should answer:
  - *are the annual maxima and minima for the period reasonable by comparison with other years?*
  - *do the stage and flow maxima/minima correspond (if not, is it explained by rating changes)?*
  - *do the ranges of the ratings (appear to) cover the extremes?*
7. Stage plots - with gaugings, stage for “base” flow, difference gauging WL and WL for gauged flow (i.e., Tideda bed plot) - for the review period plus the previous year. Both a full range plot and a partial range selected to show the presence or absence of errors at low and medium levels. Questions are:
  - *are there any out-of-character spikes, straight lines, flat peaks, etc., which are not explained in either the initial or dated comments?*
  - *are there any steps or differences between batches?*
  - *are recession shapes consistent?*
  - *do gaugings plot on?*

- *does the plot of stage for “base flow” show rating changes which are generally consistent with the height of recessions?*
8. Flow plots - for the review period plus the previous year, with gaugings plotted - of a partial range sufficient to show the presence or absence of errors at low and medium levels. Questions are:
- *are hydrograph shapes reasonable (as for stage plots above)?*
  - *do all gaugings plot on except for those which are commented?*
  - *are recession shapes and values consistent?*
  - *do ratings appear to have been consistently applied?*
9. Decade flow plots - as above, but with the previous years to make 10 years. Questions to answer are:
- *are there any trends evident which are not explainable from the comments, and which could be due to errors such as change of recorder zero, recorder location, archiving errors, etc.?*
10. Rating plots - for the review period plus the previous rating, with gaugings, both full range and a partial range up to 1 or 2 times mean flow. Use colour if available. (In some circumstances Gen-a-Rate may give the best presentation - but won't run in script file at present).
- *are the rating shapes consistent?*
  - *if ratings cross, is this explainable in the comments (by cross-section or reach shape changes), or could a rating change have been missed?*
  - *does the top end treatment of merging or parallel ratings appear appropriate?*
  - *are gaugings which plot off the curve by >8% explained in the comments?*
  - *no rating changes appear to have been missed?*
  - *extrapolations of curves appear realistic?*
  - *extrapolations have a reasonable basis (e.g. V-A extension, in comments)?*
11. Comparative plots - it is left to the review compiler's discretion to choose and present a method of comparing two or more stations (flow, rainfall, gate openings, independent backup recorder, same/nearby catchment, or ?) to demonstrate that there are no obvious anomalies in the record. Statistical techniques such as double mass curves, direct comparisons of parameters, or relationships between parameters (e.g. addition, subtraction, correlation) may be employed. (It is intended that this area of the review be developed further in the future, and for the present less than ideal techniques may need to be used.) Notes will need to be provided to provide the auditor with an understanding of the comparison, to answer the following questions:
- *does the comparison method seem valid, or at least, is this the best available at present?*
  - *are there any apparent anomalies which need investigation, having regard to the notes explaining the relationship?*
  - *do flood events, recessions, control events, etc., correspond as applicable?*
  - *are there any events which appear to be missing without explanation?*
  - *are there any other unexplained anomalies?*
12. A copy of the latest station inspection form, to answer:
- *are there any issues which could affect the data accuracy which remain unresolved?*
  - *are there any important features which have not been commented on?*
  - *is the station apparently being operated effectively in the field?*
  - *has an inspection been done within the last two (calendar) years (to enable the above two questions to be answered)?*