

IID WATER INFORMATION SYSTEM -- IRRIGATION DATABASE  
PRINCIPLES AND MANAGEMENT

Bryan Thoreson<sup>1</sup>

Anisa Divine<sup>2</sup>

Mike Archer<sup>3</sup>

ABSTRACT

The Imperial Irrigation District (IID) and the Metropolitan Water District of Southern California (MWD) entered into a pioneering water transfer agreement in 1988. MWD was to finance water conservation programs in the IID in return for the transfer of the conserved water volume each year for 35 years. A requirement of the agreement was that the water conserved be verified.

In 1993, the authors spearheaded collection of conservation verification data using the WCC SCADA system. As measurement structures and data collection equipment were installed and calibrated, data collection began. Initial examination of the data showed that certain types of data errors occurred repeatedly. A Fortran program, developed to incorporate these checks, was used in conjunction with spreadsheets to achieve a high level of consistent, automated data quality control. Meanwhile, other IID departments were collecting data using Stevens Charts and data loggers, with the logger data being processed in a spreadsheet. Manual quality control was performed on both types of data.

The IID/MWD program began developing a Water Information System (WIS) incorporating daily quality control operations and a data storage warehouse function for site-specific, time-series data related to the flow of water through IID's irrigation and drainage system. The WIS also provides an audit trail of the data elements as they flow through the quality control operation. Once a month, graphs of the data are printed, checked for final quality control, and archived as hard copy. A *Processed Flow Data* document is published annually, and a *Users Manual* is updated regularly to be current with procedural changes.

In this paper, the authors describe the principles they have incorporated and the lessons they have learned during the process of setting up the irrigation flow database. Primary among these insights are: 1) the value of keeping all data in a centralized database eliminating duplication and ensuring that all analysts use the same data, and 2) keeping all data entry and manual data quality control decentralized so that data quality control is done by those who know the data best.

---

<sup>1</sup> Project Engineer, Davids Engineering, 1105 Kennedy Place, Suite 5, Davis, CA 95616

<sup>2</sup> Water Resources Planner, Imperial Irrigation District, P.O. Box 937, Imperial, CA 92251

<sup>3</sup> Senior Engineer, Murray, Burns and Kienlen Consulting Civil Engineers, 1616 29<sup>th</sup> Street, Suite 300, Sacramento, CA 95816

## INTRODUCTION

The Imperial Irrigation District's (IID) data collection program has grown greatly in the last 20 years. Data collection has tended to be the result of specific programs often with little or no coordination between new and existing programs. Data collection programs have proliferated as water supplies remain more or less constant, while increasing demand for water results in demand for better water management. In addition to the often overlapping data needs of various programs, similar data are also needed for operations. A common result is the collection of duplicate data, likely processed differently, resulting in slightly different historical records for the same site stored with different units.

During the same time span, data collection technology has undergone significant change. The preferred method has shifted from analog Stevens Recorder Charts to digital data loggers and is moving to radio telemetry. The advent of personal computers and office networks has provided the technology to easily share data and allows many people in disparate locations access to a single data source.

At the Imperial Irrigation District, the verification data requirement of the IID/MWD water conservation program has been the catalyst bringing the new technology together with the need for more data. Water conservation verification requires many types of data from disparate sources, and provides needed funds to drive data integration. Thus, the IID's Water Information System (WIS), an integrated irrigation database, was born. The WIS contains information important not only for the conservation verification program, but also for improved operation, maintenance and management of IID's irrigation distribution system and for future planning activities. For enhanced IID on-farm water management assistance to farmers, the WIS may one day include weather, evapotranspiration (ET), and tailwater return system (TRS) operational data.

## BACKGROUND

IID's first Stevens Recorder was installed in 1914, soon followed by others, until by 1980, some 75 Recorders were in service. The data collected thereby was archived on Stevens Recorder Charts in analog format. In the mid-1980's, IID also began using data loggers to collect data in electronic format, with the data stored on a PC in ASCII text file and spreadsheet format. In addition, the IID has a mainframe, AS400 system which was developed to handle IID's accounting and water order functions. The responsibility for each of these systems lies with a different organizational unit in IID, with data developed to serve a variety of functions.

Staff processed Stevens Charts weekly. Notes were entered to indicate site conditions that affected the data. These entries indicated times when the site was out of operation, when a storm had occurred, when the pen had run out of ink or the chart slipped off the cylinder, and so forth. Often the zanjero (ditch tender) left notes regarding changes at the site, particularly at sites where grade boards were used as the measurement structure. Sometimes these movements could be determined by staff, even without notes from the zanjero. Finally, a flow was recorded on the chart for each day based on each day's average water surface level (head). As of this writing, only a few Stevens Recorders remain in operation in the IID, being replaced by data loggers or SCADA.

Data loggers are in use throughout the IID. Staff collects the packs on a regular weekly or biweekly schedule. Before the development of the WIS, the data were transferred to a PC, and processed in spreadsheets. Processing mainly consisted of adjustment for grade boards, based on notes left at the site by the zanjero and the judgement of the processor, and entry of zeros for any period of missing record (gap). Logger output consisted of hourly flow calculated as an average of four 15-minute water surface levels. Plots of the water levels and flow were printed and distributed weekly.

The AS400, in addition to accounting functions, is used to record water transactions, including orders, changes in orders, and deliveries to each farm delivery gate; cropping patterns and acreage; and other data. These entries are based on IID's accounting procedures as well as zanjero field reports.

Thus, by the late-1980's, IID had data sets in three different formats with each set collected for different purposes, processed by different persons using different methods, and stored in different locations.

## ANALYSIS POTENTIAL

### Organizational History

IID's data collection, processing, and management system worked well for operational (water movement) and financial (water accounting and payment) needs. However, with the advent of an elevated water conservation ethic, other needs arose, for example planning and implementing water conservation programs, and verifying their effectiveness. At the same time, technological advances allowed the District to implement systems that would facilitate these new needs.

One project developed under the IID/MWD program, the automation of many of IID's main canal structures with a SCADA system, uses radio telemetry from the new Water Control Center (WCC). The SCADA system, which depends on the

transmission of data, allows IID to collect data from many sites, both those designed to conserve water in IID for transfer to MWD and those which IID developed on its own.

Two types of IID/MWD water conservation projects have large data collection and processing requirements. System-level projects (mainly interceptors) are under the jurisdiction of the Water Resources Unit (WRU), while the Irrigation Management Unit (IMU) is in charge of farm-level programs (mainly tailwater return systems).<sup>1</sup> Once the data collection began, the data was reviewed and the need for quality control (QC) and data processing was evident. The WRU staff, under the direction of the CVC, developed a data QC and processing procedure.<sup>2</sup> Meanwhile, the IMU staff developed a method of data QC and processing appropriate for the on-farm systems.

Furthermore, processing Stevens Charts to develop a historic record compatible with the level of accuracy available from the electronic data, led to the realization that having data in electronic form greatly increases its usefulness for analysis. Electronic form facilitates many different analyses with a minimum of additional processing time. This was a major factor driving the replacement of nearly all Stevens Recorders throughout the District with data loggers. This logger data is processed by the Hydrography Unit staff, which has recently been placed under the direction of the WRU.

Thus, IID has largely moved from an analog, paper-based data collection, processing and warehousing procedure, to one where the system is almost entirely electronic, with charts printed monthly to maintain a written data archive. However, daily log sheets, which record operational flows as reported by the zanjeros, are still in use, with entries hand-written by dispatchers in the WCC.

Nearly all WRU data is SCADA-based, while both SCADA and logger data are used by IMU. In addition, due to the water ordering and accounting systems of the mainframe AS400, a large amount of farm-level data are available. The result is the availability of a lot of data which may or may not have been recorded, processed, or calculated using the same procedure; and which may or may not be in compatible format for analytical processing. In 1996, with the advent of data QC, processing and warehousing in IID's WIS, this disparity began to change; and a process of data integration began.

### Transaction Processing versus Analytical Processing

---

<sup>1</sup> These programs are described in the accompanying paper: On Farm Irrigation Water Measurement and Evaluation by David Bradshaw, Imperial Irrigation District, Imperial, CA.

<sup>2</sup> The CVC mandate and the QC and data processing procedure are described in the accompanying paper: Irrigation Flow Data Collection, Quality Control and Site Monitoring by Michael C. Archer, Anisa Joy Divine and Bryan P. Thoreson.

Many irrigation districts, including IID, have had an on-line transaction processing (OLTP) system to process water orders and billing. IID's system uses a mainframe computer (AS400) to store data in "flat files." This system has almost 12 years worth of information about water ordered and delivered along with crop acreage and other data. Although this historical data is valuable for planning and analysis, the structure of the OLTP system does not allow optimum performance for analysis. This, along with the fact that the peak use of the OLTP system for order processing (late morning to late afternoon) coincides with the peak use time for analysis, makes planning and analysis functions difficult using this system.

These problems have led to emergence of on-line analytical processing (OLAP) and data warehousing. In these systems, historical data from the OLTP system is stored on another system dedicated to planning and analysis and structured as a relational data warehouse for optimum query performance. Historical data can be moved from the OLTP system into a different structure on the OLAP system in batch routines scheduled to run during the night at predefined intervals, i.e. daily, weekly, monthly. The interval selected depends on the need for current data. OLAP systems are designed to store millions and millions of rows of historical information and still allow users to run queries that return results in a few minutes.

#### DATA INTEGRATION

IID is collecting, processing and warehousing 15-minute time series data at over 150 SCADA sites. Another 70 or so logger sites provide hourly time series data. The SCADA and logger data are processed to provide hourly and daily flow data for accounting and analysis. The hourly and daily flow data are available in their processed form on the WIS. In addition, the raw data are stored and can be viewed, but not modified, by staff. Figure 1 shows the main data categories that are being incorporated into the WIS.

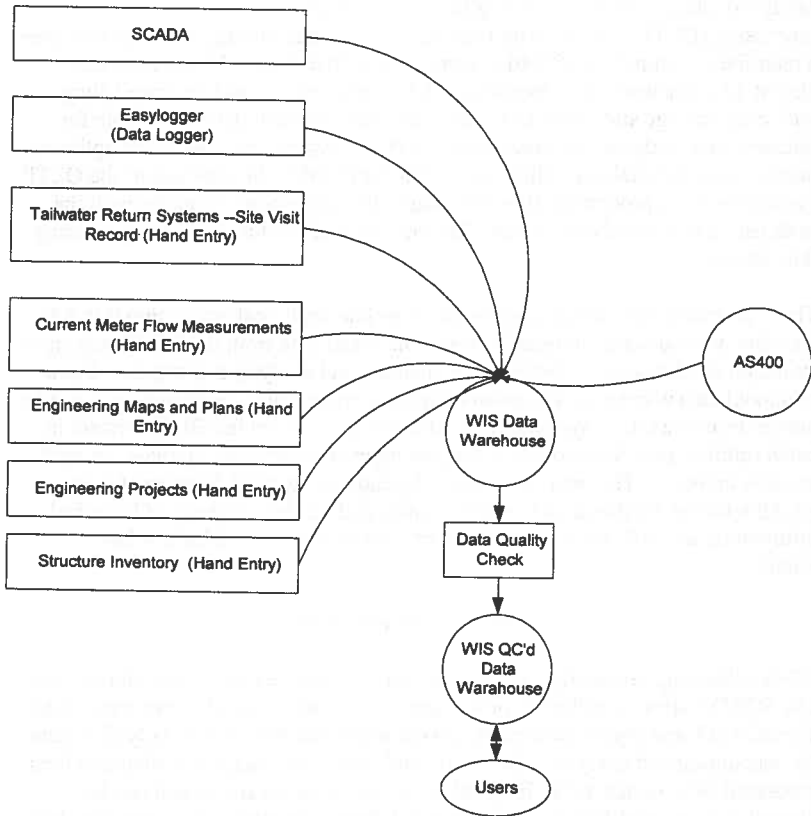


Fig. 1. WIS Conceptual Data Flow Model

Given that the WIS is a LAN-based system, manual quality control has remained decentralized in the areas of responsibility and interest: WRU, IMU, and Hydrography. Other collectors of data may join the system, with the ability to develop their own data QC and processing procedures. Each unit collaborates with WIS programmers to develop a QC processing routine for its data. Next, forms and sometimes graphs are developed that allow the final manual QC processing step to be completed by those who know the data best. Thus, even though the data is processed, warehoused and accessible from a central location, the final QC processing step remains decentralized. These final QC activities are usually carried out once a month before the summary process calculating hourly and daily averages for the previous month is run. The process for SCADA and logger sites is similar. Figure 2 diagrams the quality control process for data from SCADA sites.

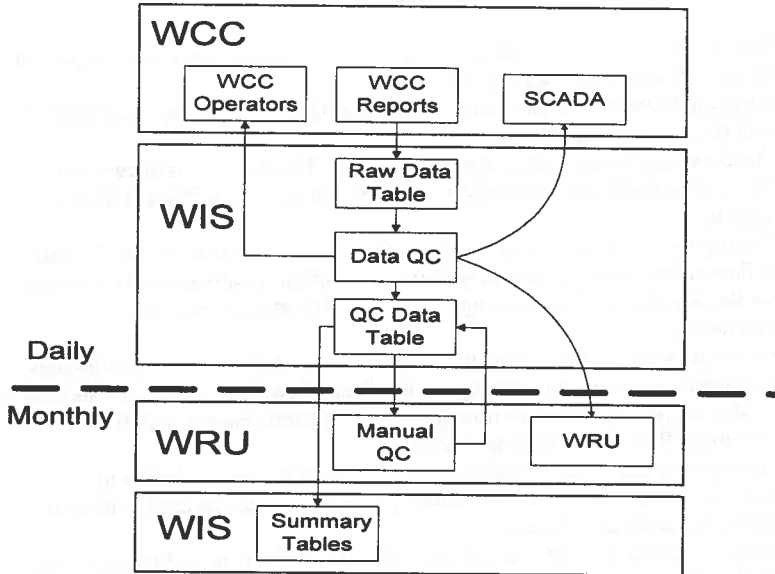


Fig. 2. SCADA Data Quality Control Process

For SCADA sites, the raw 15-minute time series data are processed daily, shortly after midnight. Only the current year of processed 15-minute data remain available on the WIS. Previous years' data are archived to CD-ROM for permanent storage. For logger sites, the processing occurs after the packs have been collected and entered into the system by Hydrography staff.

Several advantages accrue from this integration. One is that collection and processing of duplicate data is eliminated. In addition, the level of data QC and type of processing is well documented. Finally, analysts and planners throughout the District have access to the same data – which are becoming available at their PCs in forms that are increasingly easy to use.

#### FLOW DATABASE DEVELOPMENT PRINCIPLES

During the design, development and evolution of the WIS, a list of important principles to consider during design and development of irrigation databases that will store many years of data was collected. These principles will help ensure that

data can be accessed quickly and easily for analysis and planning purposes. These principles are listed and discussed below.

1. Store all data in a centralized database to eliminate duplication and ensure that all analysts are using the same data.
2. Maintain decentralized data entry and manual QC. Thus, those most familiar with the data are responsible for QC.
3. One flow equals one unique physical location. This allows for integration with a geographic information system (GIS) and facilitates changes in data collection.
4. Classify flow sites according to type of flow measurement at the site for ease of flow calculations. Calculating flow as part of the quality control procedure has the advantage of maintaining a history of site rating curves and parameters.
5. For use in water balance calculations, sites can be classified as: a) spill--flow from canals to drains, b) discharge or interface--flow from canals to reservoirs or other canals, c) headings--flow from main to lateral canals, and d) deliveries--flow from canals to farmers.
6. Classify sites according to "destination" and "origin" pools of water to facilitate volume balance calculations. These classifications can be made at both micro and macro levels.
7. Apply QC codes in a defined priority order, thus a "bad" record receives only the first QC code that applies to it. The remaining QC checks are skipped, once a record has been declared "bad."
8. Classify QC codes according to general conditions. For example, three possible QC code classifications result from: a) QC program checks, b) manual data review and c) variables affecting the flow calculation.
9. Make QC codes general, not referencing specific sites or values. Memos can be written and stored in the database, or site books to describe specific situations that require more explanation. If QC codes are written referring to specific sites or values, the number of QC codes quickly proliferates when the database covers many years.
10. Users who manually enter QC codes have their user identification codes and the date entered in fields on each record they modify. This provides a complete audit trail of the QC process.
11. Store data in only one location, and enter data using lists of values (LOVs) as much as possible to prevent spelling errors. LOVs allow users to enter data based on data already stored in the database so users do not have to type in values.
12. Use a constant flow data collection interval. The interval depends on how often the flow changes. Base the interval on the sites with the greatest data variance – most likely flows from the end of lateral to drains, for which a 15-minute interval has been chosen. Using a constant time step allows accurate hourly, daily and monthly averages to be calculated using relational database functions. Whereas, direct use of the average function to determine monthly



and daily averages gives incorrect results when a variable time step is used. A constant time step also makes it easier to predict how much storage space will be needed. Variable time steps may decrease, or increase, storage requirements depending on the parameters used to define when a new value is recorded.

13. Create summary tables with daily and hourly flow averages to increase the speed of access to daily and hourly data that users will need more frequently.
14. Estimate all missing flows using carefully defined procedures depending on the amount of missing data and the timeliness of the required data (flow with estimates required immediately, after one week, or after one year).
15. Use data warehousing principles (Kimball, 1996) and table structures to keep the number of tables to a minimum and relationships between tables simple. Time series flow and level data are stored in a fact table linked to dimensional tables with many, many fewer records that describe the sites, codes and time periods that cannot be easily calculated with standard date functions. The links to these smaller tables are used in queries to limit the number of records that must be accessed in the larger time series data table. These techniques increase query performance.

### Documentation

Two volumes have been developed documenting the system. One volume, WIS Warehouse Project System Reference Manual, documents the database structures and related procedures. The second volume, WIS Warehouse Project System Users Manual, documents the user menus. These documents have the added advantage of documenting the reports produced and various procedures used to account for water.

## APPLICATIONS

At present, five applications are available to Water Resources Unit staff. These are QC Management, QC Update, QCD Reports, a New Site Form and an Inventory Form. These applications allow any user with rights to perform QC management and update functions, to enter new sites, to access the data in report form, and to enter IID operational site inventory data. The screen for the QC Update menu is shown in Fig. 3 as an example.

A logger application is available to Hydrography Unit staff for loading logger data and performing QC procedures, plotting graphs, and running an annual report. The Current Metering Menu can be accessed by the Chief Hydrographer to enter data and print graphs of rating curves along with the metered points.

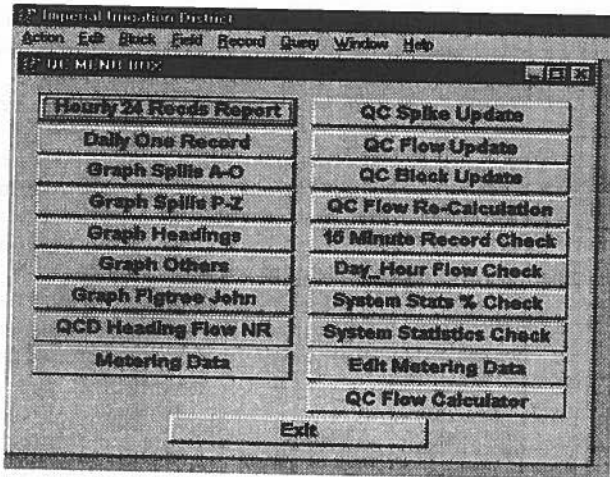


Fig. 3. The QC Update Menu

#### Daily System, Site and Data Monitoring

WRU staff checks for WIS messages each day to ensure that the data QC procedure has run and that the sites and SCADA systems are operating as expected. In the event that a problem is found, the proper individual is notified so that the problem can be rectified. In addition, WRU staff can run weekly reports to determine the amount of spill along each interceptor and the amount of water discharged into the interceptor reservoirs.

The current metering application allows WRU to calibrate the rating curves for sites where this is needed. This application allows for current meter measurements to be accessed quickly and in a useful format for rating curve development. A graphical application was developed to allow staff to quickly evaluate rating curves versus the current meter measurements (Fig. 4).

#### Monthly Quality Control

Once a month, the graph applications are used to print a graph for each site. Each graph is inspected for possible quality control requirements, including such things as sudden, large flow rate changes; erroneous flow data associated with a flat overshoot gate; and gaps in the heading data. In addition, the record is checked to ensure that the correct number of records are contained in the data set for the month. The data is then manipulated using various other applications to complete the quality control of the 15-minute time series data before the running of the summary process on the second Sunday of the month to calculate hourly and daily flow averages.

Redwood Canal Heading BCW  
Rating Curve No. 19951103 (WIS Active Date)

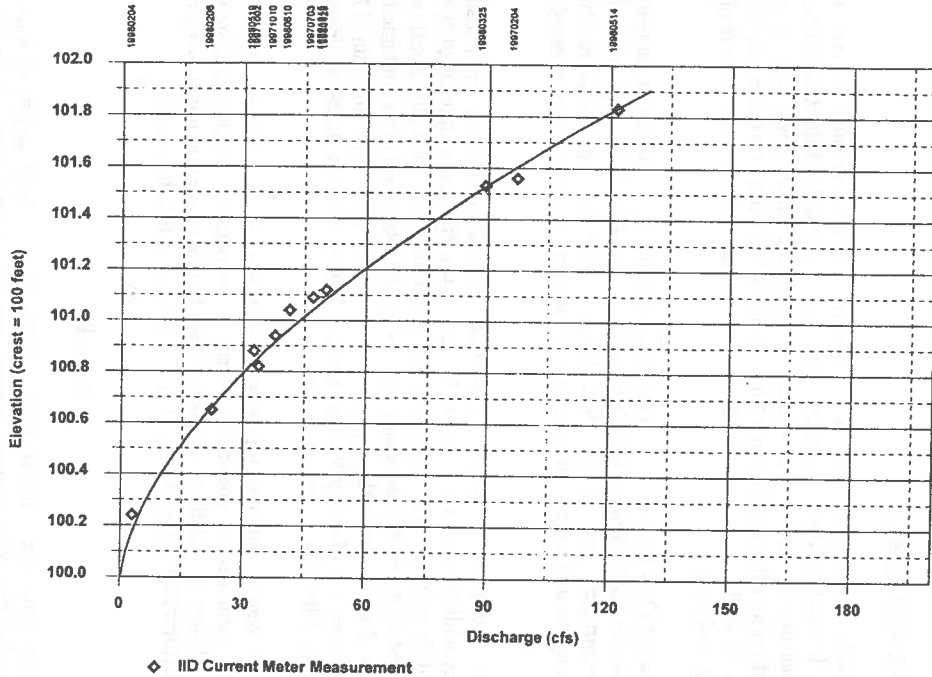


Fig. 4. Rating Curve and Current Meter Measurements Graph

Date Graphed: 07/17/98

Prepared by: WIS, IID

Finally, each graph is archived in a "Site Book." This is a book for each site consisting of pertinent information about the site, and the monthly graphs. In this form, IID is able to keep a "hard" copy of the data.

### Processed Flow Reports

Mean Daily Flow reports (Fig. 5) can be printed for all WIS sites, SCADA and logger. In addition, daily averages of digitized Stevens Recorder Chart data, which has been loaded into the system, can also be printed in this form. Tables of Mean Monthly Flows in CFS and Monthly Flow Volumes in AF can also be printed. In addition, once a year the WIS programmer prints a set of plots of the Monthly Volumes.

A Processed Flow Data document is compiled annually. This report contains the data used for verification of water conserved by six IID/MWD projects. Included in this report are an introduction, including site name conventions, abbreviations and acronyms, and tables which graphically indicate the period of processed record.

The next section of the Processed Flow Data document contains a detailed site summary table to assist the reader to locate original files, as well as identify the nature of the site and any notes that might affect the data. An alphabetical index is provided so the reader can easily locate the data tables and plots contained in the report. Finally, for each project, Notes, a Project Area Map, the Annual Mean Monthly Flow and Monthly Volume tables, Monthly Flow in Acre-Foot Plots, and Mean Daily Flow sheets are presented.

This report was initially developed using databases and spreadsheets. This procedure contained the possibility of error as the data was transferred from place to place. Using the relational database, these functions are not only much more easily performed, but the chance of error has been reduced, as well.

## FUTURE

Many applications have already been developed for the WIS, and many more are envisioned as additional data are incorporated into the system and development continues. Some future uses include: automated water balances and regular reports.

### Automated Water Balance

Addition of a few sites and minor further development work is required to allow monthly and yearly water balances to be calculated as a WIS report. This will allow regular analysis and tracking of system performance ratios and trends.

Imperial Irrigation District  
Cat Lateral Spill  
03CAT\_031\_S  
Mean Daily Flow in Cubic Feet per Second

YEAR: 1997

DAY	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
02	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
04	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
06	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
07	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
08	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
09	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0e	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3*	0.0	0.0
19	0.0e	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.3e	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2*	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0e	0.0	0.2	0.0	0.0
23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0e	0.0	2.0	0.0	0.0
24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0e	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0e	0.0
28	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.0	3.4	0.0	0.0
Mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.0	2.0	0.0	0.0
AC-FT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	2.0	0.7	0.0	0.0

e - 100% of daily volume estimated  
\* - 50% or more of daily volume estimated

Mean Flow = 0.0 cfs  
Total Volume = 9.1 ac-ft

Notes: Day begins at midnight (0000 hrs).  
Estimated flow for a missing record gap is computed as the average flow of the records preceding and following the gap equal in number to the missing records in the gap.

Date Printed: 03/11/98

Prepared by: WIS/IID

Fig. 5. Example Mean Daily Flow Site Report

### Quality Data Available for Regular Reports

Additional regular reports could be run from the WIS. Only one or two sites remain to be added to the WIS to complete many of these reports.

Even if no more sites are added to the WIS, it is much easier to produce reports and has improved documentation of data collection and quality control.

### CONCLUSION

The IID WIS is currently loading and quality controlling over 150 SCADA sites, a total of 14,880 data records (rows) each day. Following quality control, a detailed report is generated listing the problems at any site with greater than five percent of the total records for the day being "bad" is provided to the WRU. Next, a report is written to the operators in WCC indicating the daily flow and level averages at over 30 sites.

Data for around 70 logger sites are loaded into the WIS. Quality control is then performed, and the WIS-generated reports are distributed to IID staff. Once a year, a processed flow report is published reporting daily and monthly summaries of selected flow sites. The current meter graph is used regularly to check the rating curves at various sites.

The WIS has improved the ability of IID staff to access data, perform data quality control and assurance, and account for water in the IID's distribution system. The WIS has also increased the speed of data access and improved the confidence in the data by standardizing QC functions.

### REFERENCES

- Kimball, R. 1996. *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley & Sons, Inc. New York, NY.