FINAL REPORT AAFC-2

File Organization and Linkages for the Pilot Watershed Study Digital Database

July 1995

SSC Contract No. 01950-4-1313/01-XSE

Compilation of a Computerized Database using Data Assembled under the Pilot Watershed Study of the Soil and Water Environmental Enhancement Program (SWEEP)

Prepared by:
Andrew Couturier
358 Metcalfe St., Guelph, Ontario
N1E 4Z7

On behalf of:
Agriculture and Agri-Food Canada
Centre for Land and Biological Resources Research
Land Resource Division
70 Fountain St., Guelph, Ontario
N1H 3N6

EXECUTIVE SUMMARY

This report illustrates the organization of the Pilot Watershed Study digital database which now forms part of the broader database of the Ontario Land Resource Unit. The database is structured in a three-level hierarchy. Spatial resolution increases in the transition from the overall watershed scale, to the sub-watershed level, and finally, to the site level. The paper describes the linkages between spatial features and attribute information existing at different levels of the hierarchy. These data will be used to apply and test the IROWC methodology at the micro-basin and small watershed scales.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION AND OBJECTIVES

This report is the second in a series concerning data gathered during the course of the Pilot Watershed Study (PWS), a major research sub-program of the Soil and Water Environmental Enhancement Program (SWEEP). The initial report in this series (Couturier 1995) supplements the existing SWEEP reports by documenting the nature and quantity of data collected during the PWS available in digital format which i) have potential application to the Indicators of Risk of Water Contamination (IROWC) currently being developed by MacDonald and Spaling (1995a,b), and which ii) are appropriate for incorporation into the Ontario Land Resource Unit (OLRU) database. This paper builds on the initial report by describing the organization of the digital PWS database and the linkages among its components.

1.1 Notes on File Structures and Conversion Procedures

As noted in Couturier (1995), digital data from the PWS vary in format. The majority of the digital files delivered by BEAK and Ecologistics, the two primary contractors responsible for implementing the PWS, are in a usable format. Many originally existed as ASCII text files which, if delimited by quotation marks or commas, allow for conversion between different software packages. Unfortunately, these ASCII files were saved as standard printer files, without delimiters. In order to make these files compatible with *Info*, a significant amount of "creative" manipulation was required. However, some required so much manipulation to recover that they have not been included in the OLRU digital database. The preceding report in this series indicates which data exist in digital and print format, and which data have been incorporated into the OLRU digital database.

2. SPATIAL DATA ORGANIZATION AND LINKAGES

Spatial data and associated attribute information are organized in a three-level hierarchy (Figure 1). The **watershed** scale is the broadest, consisting of the overall region for each study area (e.g. Essex, Kettle, Pittock). Note that there are no spatial entities associated with this scale, only attribute files which apply uniformly[1] to the sub-watersheds found within the broader region. **Sub-watersheds** consist of paired test and control basins, as defined for the PWS. Associated attribute files, such as sub-basin discharge and water quality data, correspond to this level of detail. The lowest level of the hierarchy, the **site level**, comprises a variety of spatial entities and their associated attribute information, e.g. soil polygons, field boundaries.

[1]The climate data described in Couturier (1995) represent the watershed level of resolution, as they are not specific to test or control sub-basins.

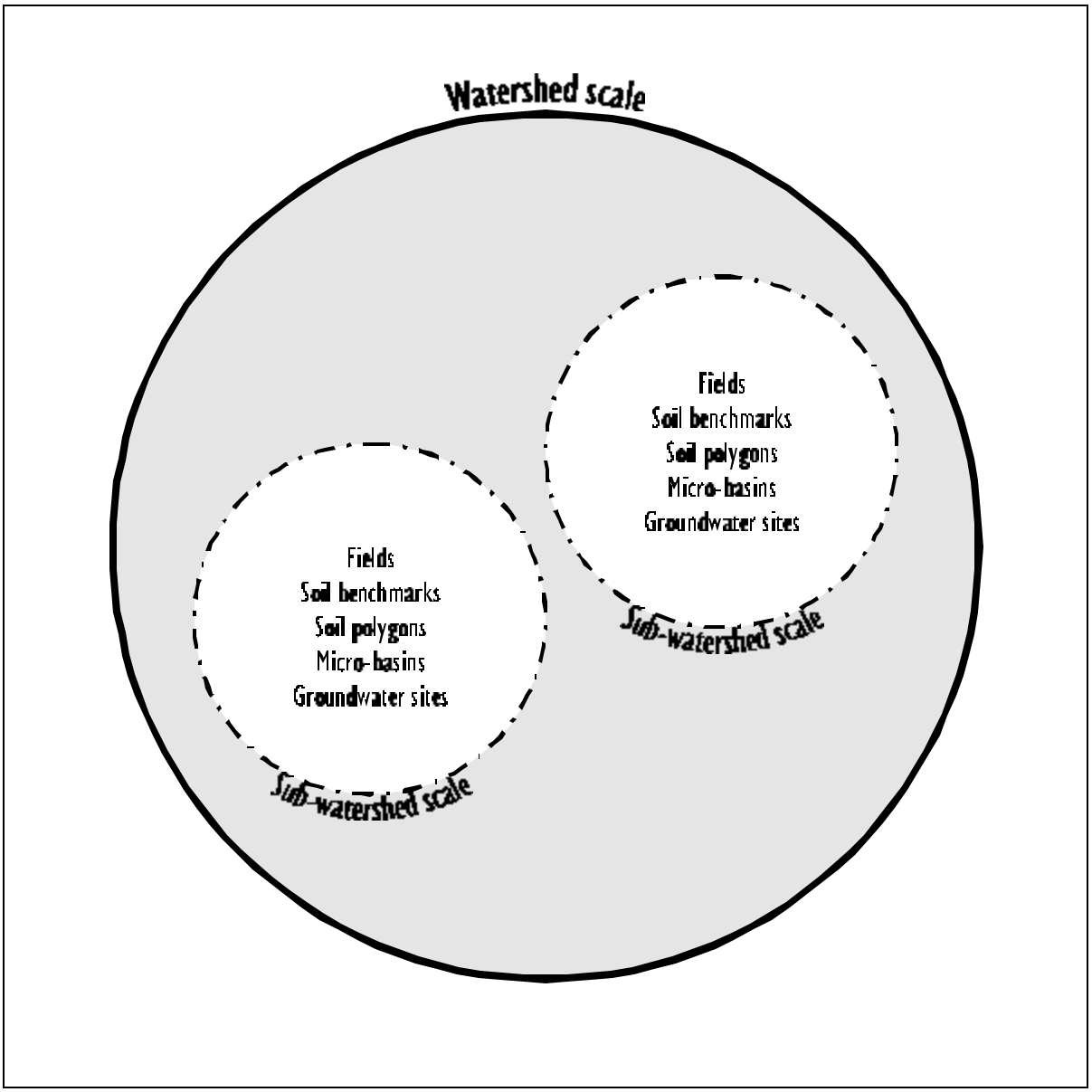

***Figure 1*. General Structure of the PWS Digital Database**

Due to its micro scale, the PWS database essentially remains independent of the structure of the existing OLRU database. However, it may be possible to register the sub-watershed boundaries to the broader OLRU database. At the time of writing, this operation had not yet been attempted. Full integration with the OLRU database has several advantages,

e.g. a better estimate of the area of farm fields and soil polygons can be determined; and, external information from OBM sheets, the SLC database, detailed county soil maps and OMAF land drainage maps can also be utilized in analyses. These linkages between the micro scale of the PWS and the broader context provided by the existing OLRU database allow for multi-scaled assessment of a variety of phenomena, including IROWC.

While Figure 1 illustrates the basic framework of the PWS digital database, the following section describes the database directory structure, the file naming conventions and the record keys which correspond to spatial entities and attribute data at the three levels of the hierarchy.

2.1 Directory Structure, File Names and Record Keys

The PWS database files are located in three **sub-directories**, whereby each sub-directory corresponds to one of the three pilot watersheds (Figure 2). All files associated with each region were placed in the corresponding sub-directory, regardless of the spatial resolution of the data contained in the files. Further disaggregation of the files into test and control groupings was deemed unnecessary; thus, in order to maintain the simplicity and manageability of the database, additional sub-directories were not created. The three-level hierarchical structure of the database noted in Figure 1 is addressed in the file naming conventions and in the record keys contained in individual files.

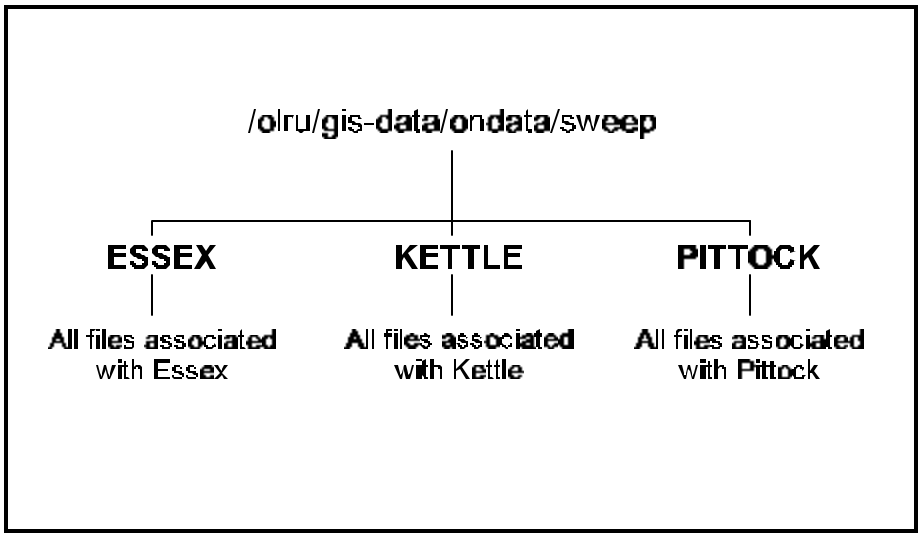

***Figure 2*. Directory Structure for the PWS Database**

Files located in the sub-directories noted above have been assigned names which provide some indication of the type of data they contain and their spatial resolution (Table 1). At the **watershed** level, data relate to each respective region (e.g. the Essex region in general) and not to specific spatial units. Attribute files are simply prefixed with ESS (Essex), KET (Kettle Creek), or PIT (Pittock) at this scale. At the **sub-watershed** scale, files are prefixed with ET (Essex Test), EC (Essex Control), KT (Kettle Test), etc., to reflect association with test or control basins. Finally, at the **site** level, file prefixes vary according to the site code or name (e.g. ECB1WQL).

Table 1. **PWS Database File Naming Conventions**

Spatial Scale	Description/File Name Prefix	Notes/Example
WATERSHED	Essex Region, Kettle Region, Pittock Region **Attribute file prefixes:** ESS, KET, PIT	**!** attribute information only at this scale (all climate data) - no corresponding spatial unit e.g. **ESSATMP** = air temperature for Essex Region
SUB-WATERSHED	Essex Test, Essex Control, Kettle Test, Kettle Control, Pittock Test, Pittock Control **Attribute file and spatial unit prefixes:** ET, EC, KT, KC, PT, PC	! data specific to test or control sub-basins (e.g. hydrology, water quality) e.g. **ETWQUAL** = water quality data for Essex Test watershed
SITE	Fields, Micro-basins, Soil polygons, Soil benchmarks, Groundwater monitoring sites **Attribute file and spatial unit prefixes:** Micro-basins = Micro-basin I.D. Groundwater = Groundwater site I.D.Fields, Soil benchmarks = ET, EC, etc. Soil Polygons (spatial) = ET, EC, etc. Soil Polygons (attribute) = ESS, KET, PIT	! micro-scale data nested within the context of broader levels e.g. **ECB1WQL** = water quality data for Essex Control Micro-basin 1

The PWS database is located in **/olru/gis-data/ondata/sweep**, where data pertaining to Essex (at all scales) are stored in the /essex sub-directory; Kettle data are housed in the /kettle sub-directory; and Pittock data are placed in the /pittock sub-directory.

At the **record level**, items containing unique identifiers provide a link between spatial entities and associated attribute information (Table 2). Depending on its position in the hierarchy, a file may contain one, two, or all of the identifier items shown in Table 2. At the watershed level, the unique identifier is contained in the NAME1 item, whereas at the sub-watershed level, it is found in the NAME2 item. In addition, the NAME1 item is also included at the sub-watershed level in order to provide a link to the watershed scale. This allows a

relation to be established between sub-watersheds (spatial features) and associated attribute information which exists at the overall watershed scale.

Finally, at the site level of resolution, the unique identifier is contained in the NAME3 item. In order to retain linkages with broader level data, the unique identifiers from these higher levels are included as items in the site level attribute and .PAT files.

Table 2. **Record Level Structure and Identifiers**

Key Fields	Description	Item Value
NAME1	Watershed Code	**ESSEX, KETTLE, PITTOCK**
NAME2	Sub-watershed Code	**ET, EC, KT, KC, PT, PC** * Plus unique identifier from above level
NAME3	Site Code	Varies with site, e.g., **ECB1, ECGL1,** etc. * Plus unique identifiers from higher levels

In addition to the unique identifiers noted above, some files require the specification of additional keys in order to select a single record. For example, climate data files contain the following items: NAME1, year, month, day, followed by 24 fields of values (e.g. 24 air temperature values). Since file structures vary according to the type and temporal resolution of the data involved, complete record layouts are provided in Reports 3 through 5 in this series.

There is some redundancy in the database design, as some information is repeated unnecessarily in the directory structure, the file naming convention, and the key fields. In hindsight, the database could have been structured more efficiently to take full advantage of its relational properties. However, the manner in which the database was constructed was largely a function of the *ad hoc* format in which the data were received from the PWS contractors. Due to the volume of data received, there was also a need to fragment the data into smaller, more manageable units.

3. SUMMARY AND CONCLUDING DIRECTIONS

This report describes the organization of the PWS digital data residing in the OLRU

database. The database is structured in a three-tiered hierarchy. Spatial resolution increases in the transition from the overall watershed scale, to the sub-watershed level, and finally, to the site level. The paper also describes the linkages between spatial features and attribute information existing at different levels of the hierarchy. These data will be used to apply and test the IROWC methodology at the micro-basin and small watershed scales. Three subsequent reports document the contents and structure of data files pertaining to the Essex, Kettle Creek, and Pittock study areas, respectively.

REFERENCES

Couturier, A. 1995. *Status and Assessment of Pilot Watershed Study Data.* Final Report AAFC-1. Guelph: Agriculture and Agri-Food Canada, Ontario Land Resource Unit.

MacDonald, K.B. and H. Spaling. 1995a. *Indicators of Risk of Water Contamination: Concepts and Principles.* Agri-Environmental Indicator Project Working Paper. Guelph: Agriculture and Agri-Food Canada, Ontario Land Reso

MacDonald, K.B. and H. Spaling. 1995b. *Indicators of Risk of Water Contamination: Methodological Development.* Agri-Environmental Indicator Project Working Paper. Guelph: Agriculture and Agri-Food Canada, Ontario Land Resource Unit.