Basic Tips for Retrieving Data from the International Stormwater BMP Database Using Microsoft Access

Introduction

This brief guide is intended to assist researchers in retrieving BMP performance study data from the International Stormwater BMP Database (BMP Database). A basic knowledge of relational databases is needed in order to use the BMP Database, and this brief "tips" guide is not intended as a tutorial for Microsoft Access. Nonetheless, these tips are written to be understandable for those with limited familiarity with database and software programming terminology, as well as those who proficiently use relational databases. The BMP Database User's Guide is a companion document to these tips that provides more detailed information on the BMP Database structure and data elements (i.e., reporting parameters or fields).

Regardless of the user's background, data retrieval should be conducted carefully and using common sense checks on the results returned. The BMP Database provides much useful information on BMP performance that can be used for many purposes; however, the BMP Database should not be viewed as a "black box" that returns absolute answers regarding BMP performance. BMP monitoring is often a messy endeavor; even the most reputable researchers can have monitoring equipment malfunctions or data entry errors that are not readily identified. One of the benefits of a large, centralized database containing studies from many researchers is that conclusions regarding BMP performance become less sensitive to minor errors and outliers associated with individual samples and studies.

Researchers should be aware that the BMP Database contains studies providing information on cutting-edge designs, as well as BMP designs that may not be optimal for water quality purposes. Alternatively, some studies may include well designed BMPs, but perhaps the BMPs have been poorly maintained or installed in manner that did not meet the design specifications. Studies of poorly performing BMPs can be just as valuable to the technical community as BMPs that perform well. (For example, bioretention cell "G1" in the BMP Database exports phosphorus due to the type of media installed in the bioretention cell. This provides valuable information on the importance of media installations closely adhering to design specifications.)

Researchers are encouraged to thoughtfully consider the objectives of their analysis as they develop queries to retrieve data. For example, research objectives could include broad generalizations regarding ranges of BMP performance where limited data screening is needed. Alternatively, more specific research objectives could be focused on the "best" performing BMPs in certain climates, BMPs in certain land uses or with certain watershed characteristics, BMPs with specific design criteria, BMP performance over a limited or broad range of storm characteristics, etc. Screening for targeted research purposes should be conducted in a manner that is objective, as opposed to selecting certain data subsets to support a pre-determined

outcome or preference for particular BMP types. The remainder of this guide focuses on practical tips for effective data retrieval from the BMP Database.

Basic Orientation to the BMP Database Structure

The BMP Database is provided in Microsoft Access 2007, which the user must own in order to use the BMP Database. The Help feature in Access can provide instruction on getting started using Microsoft Access. Basic database components (objects) include tables of data, queries used to select data of interest, forms for data entry, and reports. This tips guide focuses primarily on queries, with some background on tables.

Before writing queries to retrieve data or simply using the data tables, users should view the "Relationships Report" that provides an overview of the BMP Database structure. The "Relationships Report" is provided in the Reports section of the BMP Database. A thumbnail of the relationships report is provided in Figure 1. Primary and supporting tables are differentiated by color-coding to help users become oriented to the critical features of the BMP Database. At a minimum, users should be familiar with the "red" coded tables prior to writing queries. These tables include information on the Test Site, Watershed, Storm Event needed to link storm data together, Monitoring Stations to relate the monitoring data to the BMP, and BMP information (BMP Info). The green-coded tables provide monitoring data, and the turquoise-coded tables provide BMP-specific design information. Grey tables provide supporting information or "look-up" tables for codes (also identified by a "tbl" prefix). Dark blue coding provides supporting data such as costs, land use, monitoring agencies, etc. Understanding which tables are primary versus supplemental and viewing the BMP Database in terms of general categories of tables can make using the Database more manageable.

Figure 1. Thumbnail of Relationships Report Showing Color Coding
(see full-sized 11 x 17 version at the end of this document)
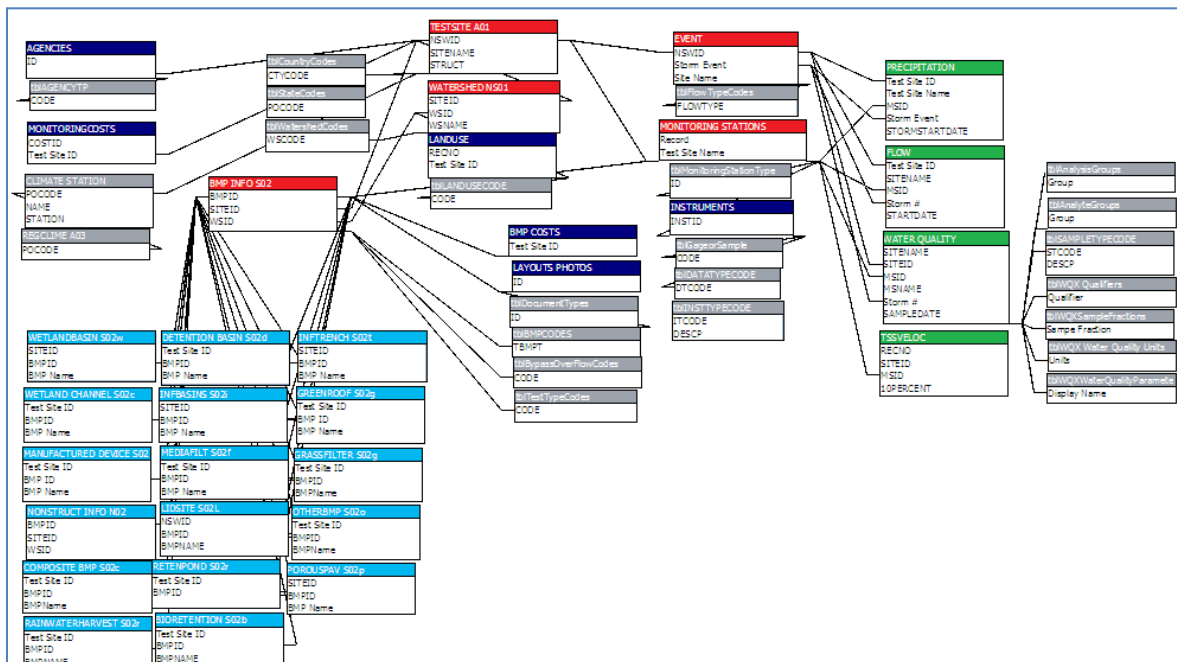
To begin writing queries, go to the "Create" tab in Access and select either the Query Wizard or Query Design options (See Figure 2 below).

**Figure 2. Create Query Tab in Microsoft Access 2007**
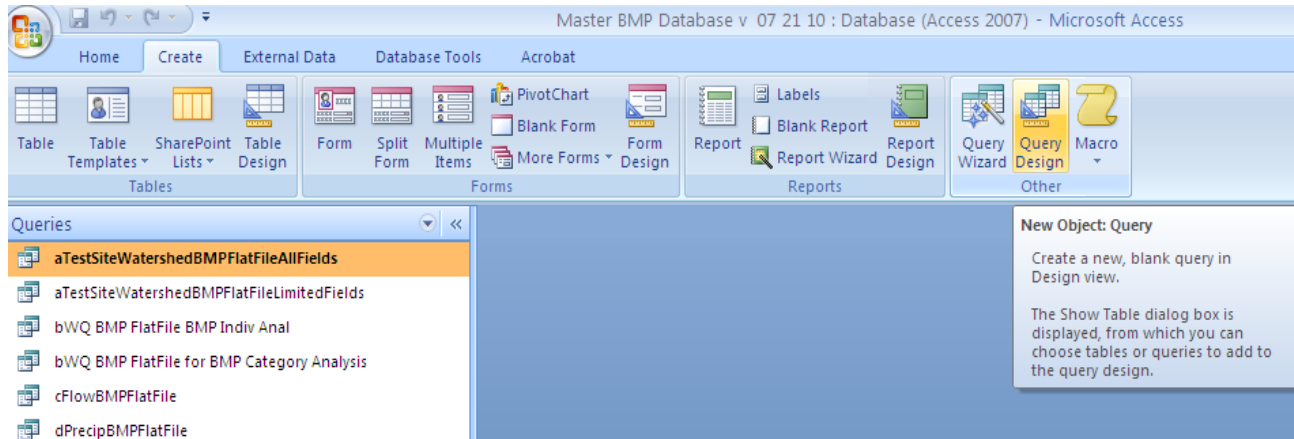
As a tool for BMP Database users, several "starter" queries have been provided in the BMP Database that establish correct linkages between various tables, essentially creating "flat files" of data that can be used as a starting point to generate more targeted queries. As shown in Figure 3, the "a" series queries create flat files of test site, watershed and general BMP information. The "b" series queries create pre-screened data sets for water quality data, with the "BMP Category" analysis representing initial screening used by the BMP Database team for category-level BMP analysis reports. The "Indiv Anal" query contains less restrictive screening for analysis of individual BMPs. (See more on this in the Monitoring Station discussion in this guide.) The "c" and "d" queries create flat files of flow and precipitation data, respectively. The "e" series queries provide queries restricted to several water quality parameter groups such as bacteria, metals, nutrients and solids. These form the basis of the initial data sets that the BMP Database Team uses to conduct analysis. Be aware that additional screening of these data sets may occur following the initial data retrieval in the starter query. Several example queries using the "Crosstab" query function in Microsoft Access are also provided. (*New examples will be added in periodic updates, as needed*.)

**Figure 3. "Starter" Queries for Retrieving BMP Database Information**
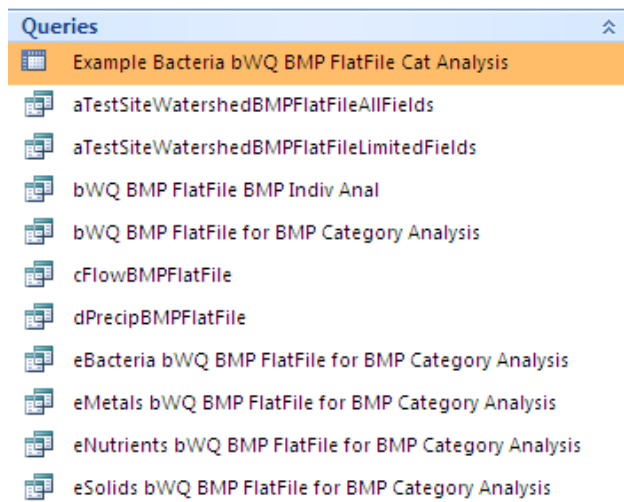
A common problem when writing queries is generation of duplicate records due to incorrect "join" properties between tables (Figure 4). For example, if you have 10 water quality records that you are pairing with precipitation event data, and your query results in 100 water quality records, you have generated duplicates. There may be an error in the "join properties" of your query or you may be using a field that does not have unique values, resulting in duplication of records. To adjust join properties between tables in a query, right-click on the line connecting the tables in the query design view. Watch for this issue when working with tables that have "one to many" relationship or "many to many" relationships. For example, one monitoring event will have many water quality records. Properly linking the "keys" (designated with a key symbol) in each data table is essential for effective queries. See Microsoft Access Help functions for more information on primary and foreign keys, and types of relationships between tables.

Figure 4. Defining Join Properties between Tables
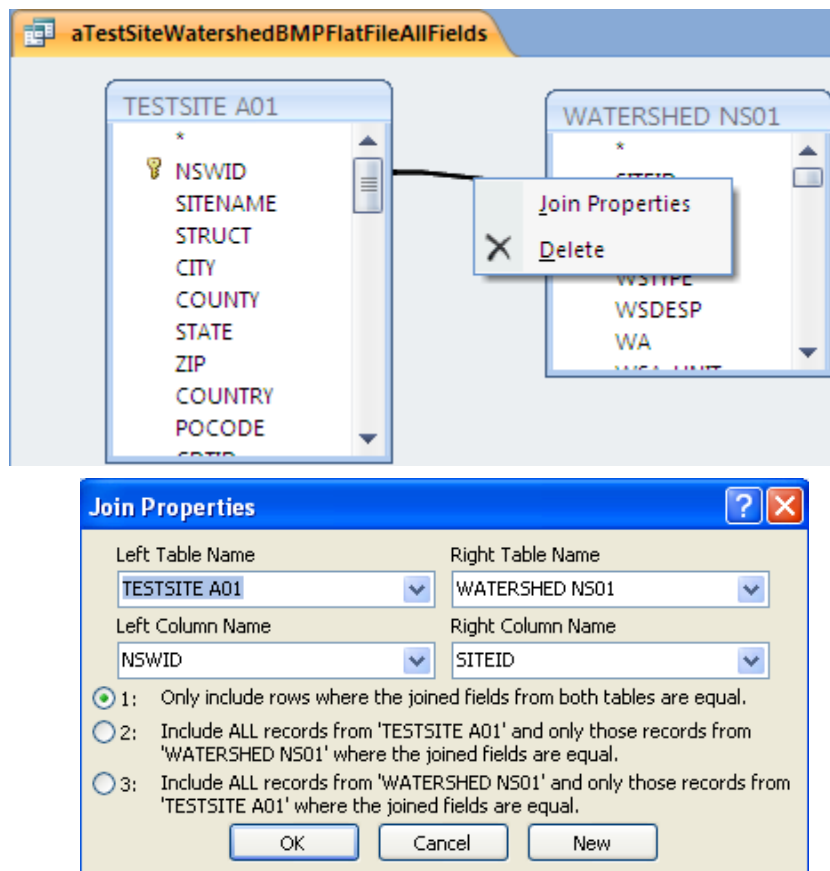
As an alternative to writing queries in Access, users may also be able to use "filter" features, similar to those in Excel, to view, copy or print data sets of interest. Filtering can be done in Table view (Figure 5), which is a user environment similar to an Excel spreadsheet.

Figure 5. Using Table Filters to Select Data of Interest
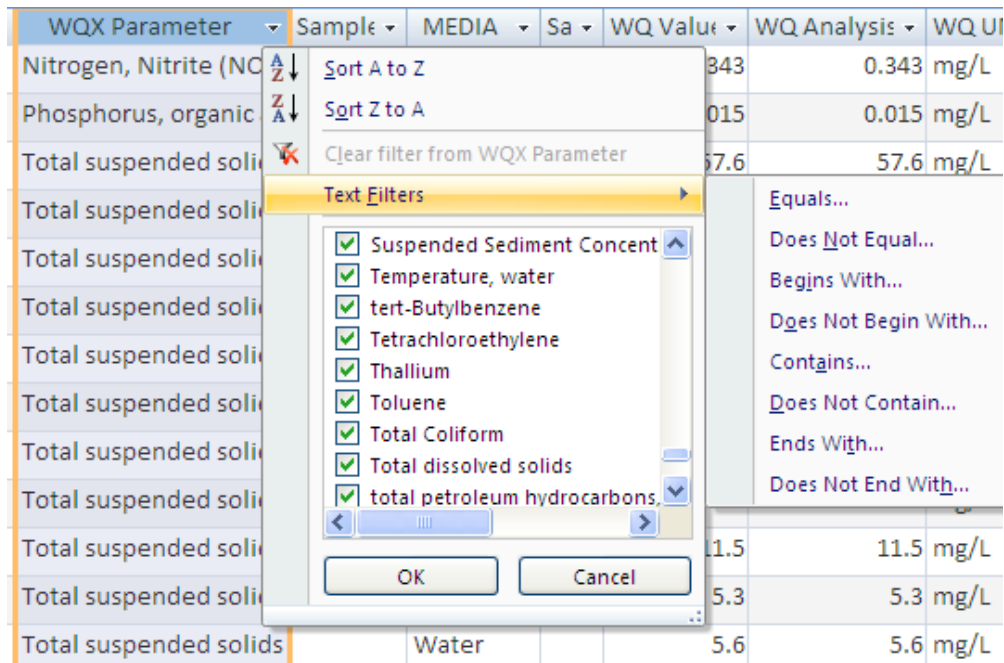
Explore and Understand Critical Tables in the BMP Database

As previously described in this guide, the "red" coded tables within the Relationships Report are fundamental to correctly using the BMP Database and are the starting place for writing most queries. Highlights of four key components of the BMP Database structure follow.

1. Monitoring Station Table

- The Monitoring Station table includes fields that enable linkages to Test Site, Watershed, BMP Info and monitoring data tables (Precipitation, Flow, Water Quality). "Foreign keys" or ID codes for these tables have been carried in the Monitoring Station table to make data retrieval simpler for users.[1] Thus, the Monitoring Station table can be a good starting point for many types of queries. Generally the field names with "ID" in their name are the keys that should be used to link tables in queries (e.g., Test Site ID, MSID, BMPID, WSID, etc.).

- The Monitoring Station table contains two screening fields that can help users "pre-screen" studies that may not be well suited for analysis at the individual BMP level or the BMP category level. Setting query criteria to exclude data sets with "No" values in the analysis

---

[1] From a software programming perspective, the BMP Database Team is aware that this is not ideal in the context of software programming standards for normalization; however, since many users of the database do not have programming backgrounds, the decision was made to "carry" these foreign keys to reduce the number of steps users must implement in order to retrieve data.

screening fields provides an initial level of data screening that may be desirable to some researchers. These fields provide a transparent basis of data screening conducted by the BMP Database Project Team. Generally, more restrictive screening is applied to Category-level analysis. The reasons for excluded data are stated in the Comments and Analysis Comments fields in the Monitoring Station Table and may include considerations such as watershed conditions atypical for urban areas (e.g., agricultural land, landfills, garage washwater), very limited numbers of storm events, designs that clearly do not include water quality components (e.g., flood control basins), and so on.

- Shared monitoring stations for BMPs in series will be entered more than one time in the monitoring station table in association with each BMP that relies upon the data collected at that station. As a result, the monitoring station relies on a "composite key" formed by BMP ID and the Monitoring Station ID. For example, when the monitoring station table is joined to the water quality data table using the Monitoring Station ID field present in both tables, the monitoring data stored one time for the shared monitoring location becomes intentionally duplicated so that it is properly associated (e.g., as inflow or outflow) with both BMPs. This is a major structural change from previous versions of the BMP Database.

- The Monitoring Station Type field in the Monitoring Station table can be used as a query criterion to restrict data sets to inflow and/or outflow or other types of data.

- When a Monitoring Station Name is identified as "FCV Inflow", this indicates that the BMP Database Project Team calculated a composite value to represent the inflow for analysis purposes because multiple inflow locations were monitored at the site. At these sites, the individual inflow monitoring locations will have a "no" in the Analysis Screening fields so that only the FCV Inflow value is retrieved when the data set is queried. A few studies also have FCV Outflows.

2. Monitoring Data Tables (Event, Precipitation, Flow and Water Quality)

- The Event table is the key to pairing inflow and outflow data for a particular storm event, as well as linking precipitation to flow and flow to water quality, etc. A simple numeric event number is provided for all of the monitoring data tables. When linking these tables together, care should be taken with regard to "join" properties in order to prevent duplication of records. The Event Table relates to the monitoring data tables based on a "composite key" formed by unique Test Site ID and Event ID.

- When joining various types of monitoring data, consider writing limited queries in small steps and linking these queries together. (The starter queries are useful for this purpose.) Be aware that some studies do not contain records for all three types of monitoring data, so the query should be written either to 1) "include all records" for the monitoring data of primary interest (e.g., water quality) and "only those records that are equal" from the other table(s), or 2) "only records where all records from both fields are equal." (See Figure 4 regarding join properties.)

- In the Water Quality table, EPA's Water Quality Exchange (WQX) nomenclature system uses two separate fields to identify the water quality parameter: WQX Parameter (e.g., lead)

and sample fraction (e.g., dissolved). Water quality queries should include both fields to prevent mixing of dissolved and total results in the query output. The supporting table "tblAnalysisGroups" is used by the Database Project Team to map these two fields into a single common name field for purposes of analysis and is incorporated in the "e" series starter queries.

- The Water Quality table contains two analysis result fields: Value and Analysis Value. The Value field is the value provided by the data provider in conjunction with a Qualifier (e.g., U for non-detect). The Analysis Value field is a calculated field calculated by the BMP Database Team for purposes of analysis that replaces "U" qualified data with one-half of the detection limit (a simple substitution method for censored data) for purposes of analysis. Bacteria data analysis values use a simple substitution method that replaces values exceeding the upper and lower quantitation limits with the reported value.

- The supporting table "tblAnalysisGroups" can be joined to the Water Quality table to retrieve categories of constituents such as metals, nutrients, solids, biological (bacteria), organics and general chemistry parameters. Similarly, the supporting table "tblAnalysisGroups" can be used to map similar parameters to a common analysis category. This table is particularly beneficial for certain nutrients that may be appropriate to analyze together (e.g., orthophosphate may be reported using various sample fractions that are assumed to be appropriate to analyze as one overall data set).

- As a general practice, users should verify that data sets retrieved report information in common units of measurement. Common units have been provided in the precipitation, flow and water quality tables.

- Avoid these common errors when writing Water Quality queries:

  o Use the Storm Event field to pair inflow and outflow data as opposed to using Sample Date. In many cases, inflows and outflows for storms may occur on different dates. Alternatively, there may be more than one storm event occurring on a particular date.

  o Specify "Media" type when writing water quality queries to avoid mixing sediment and runoff data. The BMP Database includes several different types of samples, including sediment samples. In the starter queries, "water" is specified as the media type.

  o Be aware that the Water Quality data table contains studies with grab samples, as well as EMCs. Exercise care when writing queries that retrieve both types of data, since grab samples result in multiple records per analyte per event, whereas EMCs result in one record per analyte per event.

  o The Crosstab query function in Microsoft Access has limited options for statistical outputs. When specifying a value to return in these types of queries, the user must specify whether to return first, last, average, etc. Microsoft Access does not calculate medians as a standard feature in cross-tab queries. Averages are typically not appropriate for water quality data because assumptions of normality are necessary for

the average to be an appropriate measure of central tendency. See the "ExampleBacteriaCrossTab" as an example of a Crosstab query.

- Flow data: be aware that collection and proper documentation of flow data is challenging. Several comments are provided below to help reduce misuse of flow data retrieved from the BMP Database.

    o Some flow data provided in the BMP Database clearly assume that inflow equals outflow, as evidenced by identical flow volumes. Volume reduction for these BMPs would be assumed to be negligible, and, in general, these BMPs should be excluded from analysis of BMP volume reduction.

    o Some BMP types may not have monitored inflows; this may be the case for LID designs that inherently avoid concentration of flows and rely on distributed BMPs receiving runoff via shallow flow or sheet flow. Precipitation and watershed characteristics are particularly important for characterizing hydrology for these studies. (Download the Volume Reduction technical memorandum from www.bmpdatabase.org for more information.)

    o Be aware that some studies contain flow records where outflow exceeds inflow. A wide range of explanations should be considered such as unmeasured sheet flows, groundwater inflows, monitoring error, etc. These types of unmeasured flows can also result in underestimation of volume reduction in BMPs where volume reduction is occurring.

    o Flow measurement accuracy: There a number of factors that can compromise the accuracy of flow measurements at field monitoring sites such as backwater effects (especially at higher flows), debris or sediment accumulation (debris on weir crest), unsteady/non-uniform flow conditions, seepage, and other factors. These factors may not be consistently described in data submissions and can be difficult to detect by third parties using the flow data. Until recent years, the technical community has not intensively focused on volume reduction as a BMP performance objective; therefore, users should be aware that older data sets may have less precise flow characterization. Measures of uncertainty should always be used with flow data. During 2010, the BMP Database Team will be further analyzing the flow data set and may apply additional qualifiers/recommendations for use of the flow data. For example, in some cases, flow data may be useful for proportioning sample volumes for EMCs, but not appropriate for drawing conclusions related to volume reduction. Also see the *Urban Stormwater BMP Monitoring Manual* [Geosyntec and WWE 2009] downloadable from www.bmpdatabase.org for guidance related to flow monitoring and the BMP Database "Volume Reduction Technical Memorandum" for recommended volume analysis approaches. Typically, a combination of performance metrics is less likely to result in misleading conclusions regarding volume reduction in BMPs. Users of flow data should consider uncertainty when drawing conclusions regarding volume reduction. (*This is mentioned in this paper because simple percent volume reduction queries using the BMP Database are fairly straightforward to develop; however, it is also easy to misuse the flow data set.*)

3. BMP Information

- The table "BMP Info S02" is the "mother" table for the BMP design information. This table contains general BMP design characteristics that can be linked to the various "daughter" design characteristics for each BMP category (the turquoise-coded tables in the Relationships Report).

- Use BMPID when relating BMP records to other data. Be aware that BMP Name is <u>not</u> a unique field and can result in errors combining data sets. For example, there is more than one BMP in the BMP Database called "CDS Unit;" however, the BMP IDs for these two sites are unique.

- See the "Layouts Photos" table for photos and design sketches of BMPs. Click on the object in the field to view the photos and sketches.

- The Manufactured Device BMP category contains a relatively large number of studies (>65); however, the treatment processes in these devices vary substantially. Analysis of manufactured device data should group BMPs with similar unit treatment processes together.

- Be aware that some BMP studies contain limited design information. There are a variety of reasons for this situation. Two common reasons include: older studies may not have had design information readily accessible, and studies may have been entered prior to new Low Impact Development categories being entered into the BMP Database. For example, bioretention and green roof BMPs were historically "held" in the media filter BMP type until reporting parameters for these BMP categories were developed in late 2008. The BMP Database Team is gradually working to back-fill some of these missing fields and encourages those submitting data to be as complete as possible with regard to design parameters. This information is critical to identifying the factors that lead to better BMP performance.

4. Test Site and Watershed Characteristics

Several tables can be used to support queries based on test site and watershed characteristics, including:

- Test Site could be used to retrieve data for certain states (e.g., the mid-Atlantic region), as well as to link to the supporting table ("RegClime A03") containing climate characteristics for the location. As the BMP Database has evolved over time, the Test Site ID key may be referred to several ways in various tables. NSWID, Test Site ID and SiteID are synonymous.

- Watershed (primary key is WSID) can be used to obtain a variety of information related to tributary land area characteristics such as total area, imperviousness, road and sewer system characteristics, etc.

- Land Use provides the relative percentages of the land uses present. One watershed record typically contains multiple land uses; therefore, land use is provided as a separate table (i.e., a one to many relationship is present between a watershed record and land uses).

Other: BMP Database Transition Notes for Historic Users of the BMP Database

- During 2008-2010, major revisions to the BMP Database were completed. For transparency with historic BMP Database users, some "legacy" fields used in the previous release of the BMP Database have been retained. These are identified by an "x" in front of the data element name. The two primary fields relevant to previous users of the BMP Database include the historic BMP category ("xoldTBMPT") in the BMP Info table) and the Legacy STORET water quality parameter ("xOldShortName (Legacy)"). The historic BMP type is provided because some BMP studies were moved to new BMP categories during 2008-2010, particularly manufactured devices and Low Impact Development BMPs. Legacy STORET nomenclature was a source of confusion for historic data providers. When the BMP Database was transitioned to the EPA's WQX nomenclature, published BMP study reports were cross-checked, where needed, to confirm the appropriate WQX parameter appropriate for the data set.

- "PDF ID" is a new field, originating in the BMP Gen Info table, which will form the basis for all analysis PDFs in the future. A document library containing PDFs of statistical analysis of the BMPs will be linked to these PDF IDs later in 2010. This field essentially replaces the random number generated field historically used for this purpose identified as "BMP ID."

- All BMP design information categories data now "pass through" the BMP Info table. Previous versions of the database had a bifurcated structure that separated non-structural and structural BMPs.

- The procedure defining the relationship between monitoring stations and BMPs had been significantly simplified relative to previous versions of the database.

- Linking codes such as "linktoprecip" and "linktoflow" have been abandoned and replaced with the Storm Event number. Establishing linkages between various types of flow data has been simplified.

Contact for More Information

The BMP Database Project Team is available via phone and email to answer user questions. See www.bmpdatabase.org for a complete list of project contacts, or email clary@wrightwater.com with questions.

C:\Documents and Settings\jclary\Desktop\BMP Database Working Files 07 27 09\DRAFT Tips for Retrieving Data October 2010.docx

Relationships for Master BMP Database v 3.2 August 2010

AGENCIES
ID

tblAGENCYTP
CODE

MONITORINGCOSTS
COSTID
Test Site ID

CLIMATE STATION
POCODE
NAME
STATION

REGCLIME A03
POCODE

tblCountryCodes
CTYCODE

tblStateCodes
POCODE

tblWatershedCodes
WSCODE

TESTSITE A01
NSWID
SITENAME
STRUCT

WATERSHED NS01
SITEID
WSID
WSNAME

LANDUSE
RECNO
Test Site ID

tblILANDUSECODE
CODE

EVENT
NSWID
Storm Event
Site Name

tblFlowTypeCodes
FLOWTYPE

MONITORING STATIONS
Record
Test Site Name

tblMonitoringStationType
ID

INSTRUMENTS
INSTID

tblGageorSample
CODE

tblDATATYPECODE
DTCODE

tblINSTTYPECODE
ITCODE
DESCP

PRECIPITATION
Test Site ID
Test Site Name
MSID
Storm Event
STORMSTARTDATE

FLOW
Test Site ID
SITENAME
MSID
Storm #
STARTDATE

WATER QUALITY
SITENAME
SITEID
MSID
MSNAME
Storm #
SAMPLEDATE

TSSVELOC
RECNO
SITEID
MSID
10PERCENT

tblAnalysisGroups
Group

tblAnalyteGroups
Group

tblSAMPLETYPECODE
STCODE
DESCP

tblWQX Qualifiers
Qualifier

tblWQXSampleFractions
Sampe Fraction

tblWQX Water Quality Units
Units

tblWQXWaterQualityParamete
Display Name

BMP INFO S02
BMPID
SITEID
WSID

BMP COSTS
Test Site ID

LAYOUTS PHOTOS
ID

tblDocumentTypes
ID

tblBMPCODES
TBMPT

tblBypassOverFlowCodes
CODE

tblTestTypeCodes
CODE

WETLANDBASIN S02w
SITEID
BMPID
BMP Name

WETLAND CHANNEL S02c
Test Site ID
BMPID
BMP Name

MANUFACTURED DEVICE S02
Test Site ID
BMP ID
BMP Name

NONSTRUCT INFO N02
BMPID
SITEID
WSID

COMPOSITE BMP S02c
Test Site ID
BMPID
BMPName

RAINWATERHARVEST S02r
Test Site ID
BMPID
BMPNAME

DETENTION BASIN S02d
Test Site ID
BMPID
BMP Name

INFBASINS S02i
SITEID
BMPID
BMP Name

MEDIAFILT S02f
Test Site ID
BMPID
BMP Name

LIDSITE S02L
NSWID
BMPID
BMPNAME

RETENPOND S02r
Test Site ID
BMPID

BIORETENTION S02b
Test Site ID
BMPID
BMPNAME

INFTRENCH S02t
SITEID
BMPID
BMP Name

GREENROOF S02g
Test Site ID
BMP ID
BMP Name

GRASSFILTER S02g
Test Site ID
BMPID
BMPName

OTHERBMP S02o
Test Site ID
BMPID
BMPName

POROUSPAV S02p
SITEID
BMPID
BMP Name