To: Distribution

From: Ernestine Bryant, STOrage and RETrieval (STORET), Technical Project Leader (TPL), Solutions Development Center (SDC)/Science Applications International Corporation (SAIC)

Subject: Minutes of STORET Technical Meeting

1.0 Purpose

A Meeting was held on April 2, 2004 at the SDC. The purpose of the meeting was to review and address the status of the STORET Project activities and resolve project issues.

2.0 Attendees

SDC	EPA
Ernestine Bryant	Robert King, Task Order Project Officer (TOPO)
Stephen Smith	Lee Manning, Alternate TOPO (ATOPO)
Blythe Norris	
Joseph Wilson	

3.0 Discussion Topics

The following sections detail the discussions of the project tasks and associated Software Incident Reports (SIRs).

3.1 Unit Conversion

The following decision were made regarding unit conversion.

- Volts will be the target unit for Millivolts and Volts.
- Micro-siemens per centimeter will be the target unit for the following units:
  - micro-mhos per centimeter.
  - mhos per centimeter.
  - milli-siemens per centimeter.

- micro-siemens per centimeter.
- siemens per meter.

C The previously determined unique unit “micro moles per square meter per second” will be converted to “micro moles per square meter per day”.

C The following units will be unique in that they will not be converted to any other unit:

- None.
- TU - Toxic Units, dividing 100% by the LC50 (% effluent.
- % sediment - % target/clean sed mix which produces response.
- lb/in - Tension, pounds per inch displacement.
- gpm/ft - Drawdown capacity, gallons per minute per foot.
- Langleys - Energy Intensity, Langleys (cal/sq cm).
- amps - Electrical current, amperes.

Options for implementing unit conversion were discussed. Possible areas where unit conversion could be applied are STORET v2.0 database, Central Database Update Software (CDUS), and Central Warehouse Extract, Transform, and Load (ETL). The software modifications and ramifications of implementing each option were reviewed (Attachment A). The following decisions were made:

C Proceed with implementing unit conversion in the Central Warehouse ETL. Employing unit conversion in STORET v2.0 database/data entry level would be the ideal implementation as a long term strategy.

C Refer to unit conversion columns as “standard” (e.g., STD_UNIT, STD_VALUE).

C Add the following columns to the TSRUOM table: STD_UNIT and CONVERSION_FACTOR.

C Apply a standard value of null to all textual results. The conversion routine must read the VALUE_TEXT as well as VALUE_MEASURE to properly ascertain whether or not the value is textual.

C Define the stored Result standard value as NUMBER so that conversion will not be limited to a particular precision.

Options for handling adjustment factors for temperature conversions were discussed. One option is to include a separate ADJUSTMENT_FACTOR column in the TSRUOM table. Another option is to apply a function for all temperature conversions. No final decision was determined.

Options for displaying standard result values were discussed. One option is to only show as many significant digits as the original result value. Another option is to display the value in scientific notation. No final decision was determined.

3.2 STORET Reference Table Application

B. King requested that the Project Team update the Reference Table Application to ensure the Analytical Procedure, Analytical Equipment, Citation, and Central Organization maintenance sections are Version 2.0 compliant. B. Norris presented mock-ups of a possible Visual Basic (VB) reference table application (Attachment B). Some initial comments about the proposed mock-up design:

- Use Option 1 for reference tables where there are five fields or less (e.g., Unit of Measure).
- Use Option 2 with a separate form to do data entry for reference tables with greater than five fields (e.g., Characteristic).
- Include Search Name in the Characteristic Maintenance list.
- Display IS_NUMBER and make it read-only on list boxes and data entry forms.

In the meantime, the existing COOL:Gen application will be modified in specific areas to handle maintenance of Analytical Procedure, Analytical Equipment, Citation, and Central Organization.

3.3 STORET Data Entry Module

B. King expressed his concerned about ongoing COOL:Gen support. He believes that another full generation of the COOL:Gen application will not be needed unless real data issues arise. The current plan is to stay at COOL:Gen 6.0, even though the STORET Project has a licence for Advantage:Gen 6.5. E. Bryant will look into the possibility of a dedicated server for the 6.0 Client Server Encyclopedia (CSE) if/when the Safe Drinking Water Information System (SDWIS) project upgrades to a later maintenance release.

3.4 Central Warehouse

Data transformation options for STORET were presented (see Attachment C) and the following decisions were made at the conclusion of this presentation:

- Datum Conversion options need to be researched further. Oracle Spatial's support for converting datums to the North American Datum (NAD)83 (1986) standard should be researched in more detail.

- A meeting will be held after April 12, 2004, to discuss and document the "Ad Hoc" data clean-up scripts that are currently being executed in tandem with the CDUS and ETL processes. This will support the goal of establishing configuration management control over these processes and ensure that the data transformation approaches being utilized are optimal and consistent.

Issues surrounding the current storage of User Defined Habitat Assessment Results were discussed. Future meetings will address the possible redesign of the Habitat Results area of the application. In the short term, the following changes will be made to the Habitat Results portion of the Central Warehouse to facilitate the retrieval of User Defined Habitat Assessments through the Window to My Environment interface and Station Home Page:

- A new characteristic will be added to the DI_CHARACTERISTIC table to identify User Defined Habitat Assessment Results.

- User Defined Habitat Assessment Result records in the FA_HABITAT_RESULT table will have FK_CHAR values populated with the code of the new characteristic mentioned above.

- The STATION_CHAR table will be updated to include User Defined Habitat Assessment Results.

- Stations that are only associated with User Defined Habitat Assessment Results will be flagged as visited in the FA_STATION table.

3.5 STORET Station Home Page

B. King provided technical direction to implement the activity start date range change (SIR 1555) on the STORET Station Home Page. This is a high priority task in addition to adding user-defined habitat assessments to the home page.

4.0 Action Item Summary

Number	Description	Date Issued	Status	Assignment	Date Completed
04-0001	Establish proposed schedule of deliverables not to exceed November 28, 2003.	10/09/2003	Closed	E. Bryant	10/29/2003
04-0002	Provide Unit of Measure count Structured Query Language to L.Manning.	01/08/2004	Closed	S. Smith	01/12/2004
04-0003	Provide TSRUOM description_text corrections to L. Manning.	01/08/2004	Closed	S. Smith	01/12/2004
04-0004	Determine supported Windows operating systems for Oracle 9i Release 2 Personal Edition.	01/08/2004	Closed	G. Thadkamalla	01/23/2004
04-0005	Provide TSRUOM description_text corrections to L. Manning.	01/21/2004	Closed	S. Smith	01/28/2004
04-0006	Inquire as to status of WSSERVER.JAR fix at RTP.	01/21/2004	Open	L. Manning	
04-0007	Provide TSRUOM description_text corrections to L. Manning.	02/11/2004	Closed	S. Smith	02/18/2004
04-0008	Provide histogram date range layout.	02/11/2004	Closed	L. Manning	02/17/2004
04-0009	Provide detail pertaining to official Oracle 9iAS rel 2 patch.	04/02/2004	Open	B. Norris	

ATTACHMENT A

Unit Conversion Implementation Options

Unit of Measure Conversion
Implementation Phase I
04/02/2004

Summary:

Implement target unit of measure conversion. Modify the TSRUOM table to include new columns with the target unit and conversion multipliers/factors. Modify the TSRRSULT table to include a new column for the calculated result value based on the target unit of measure.

Implementation Options:

Option 1 – Distributed STORET v2.0 database/data entry level.

- Add database columns for TSRUOM and TSRRSULT tables into STORET v2.0 tables.
- Populate STORET.TSRUOM table with new unit conversion columns.
- Include new TSRUOM in a DBFIX for user community.
- Provide a script to populate new TSRRSULT calculated result value for existing data.
- Provide means of automatically populating new result value whenever user adds new or changes existing result values in STORET v2.0. Do not alter the online application(s). Use a database trigger (insert or update) so that the target unit result value is populated regardless of WHERE the result value is set (e.g., STORET data entry application, SIMM, SQL*Plus, etc.).
- If population is mandatory, will need to run Option 2 for cases in which the above is not applied.

Option 2 – CDUS implementation.

- Add database columns for TSRUOM and TSRRSULT tables into STORET v2.0 centrally maintained database tables.
- Populate central STORET.TSRUOM table with new unit conversion columns.
- Include new TSRUOM in a DBFIX for user community.
- Populate target Result value for legacy/production data (one time only).
- Populate target Result value for newly submitted data on a regular basis.

Option 3 – Central Warehouse implementation.

- Include target result and unit columns in Result fact tables (already done).
- Populate STORET.TSRUOM table with new unit conversion columns.

- Include new TSRUOM in a DBFIX for user community.
- Populate new Result columns during ETL process.

Functions Performed:

- Normalized table structure script (Options 1 and 2 only).
- Target Result Value population script (global check).
- TSRUOM table population script.
- Database trigger (Option 1 only).

Database Scripts:

TSRUOM: Add columns to STORET.TSRUOM table (i.e., TARGET_UNIT, CONVERSION_FACTOR).

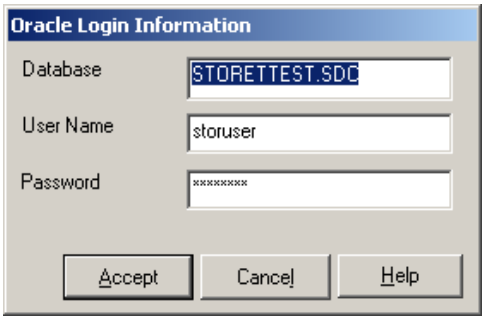
TSRRSULT: Add column to STORET1.TSRRSULT table (e.g., STD_VALUE) as NUMBER(31,16) (at a minimum). Column is not mandatory.

Database Trigger script (Option 1 only): Add script to create database trigger(s) that would perform a row-level update of the new STD_VALUE_MEASURE column in the TSRRSULT table after insert or update of VALUE_MEASURE column. It would first need to read the associated UOM to identify the conversion factor to apply.

ATTACHMENT B

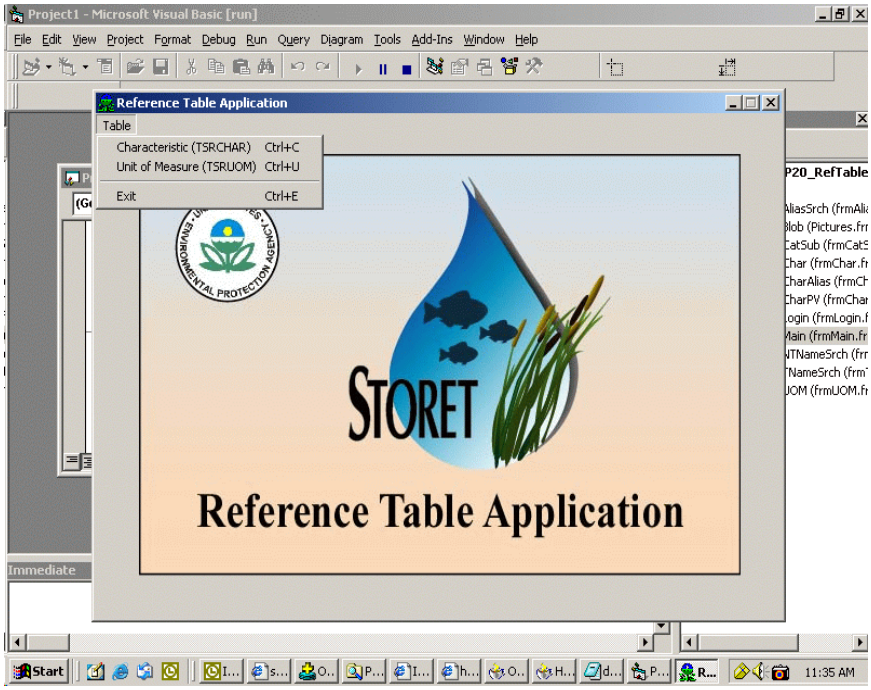
Reference Table Mock-ups

Oracle Login:

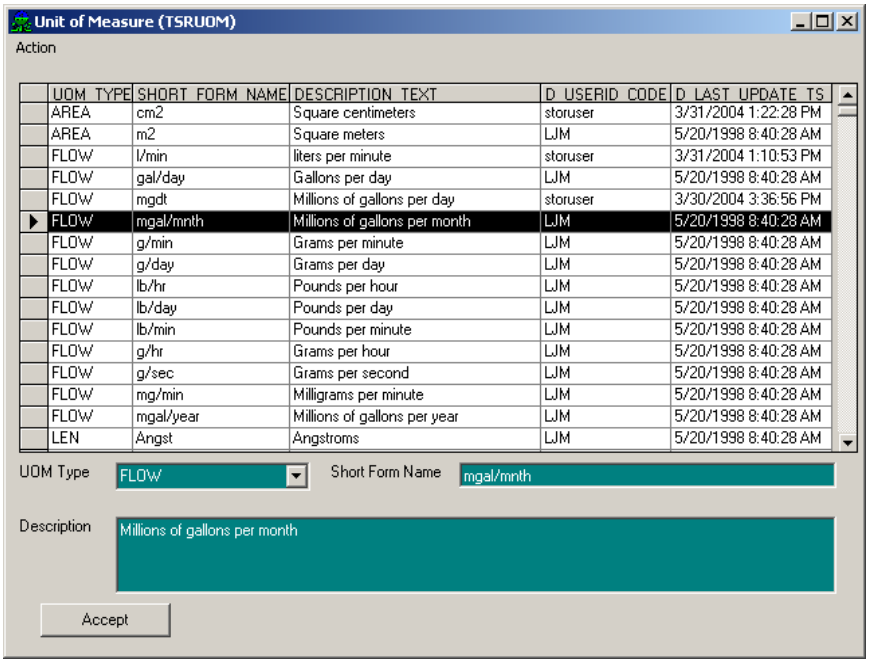


Main Menu:

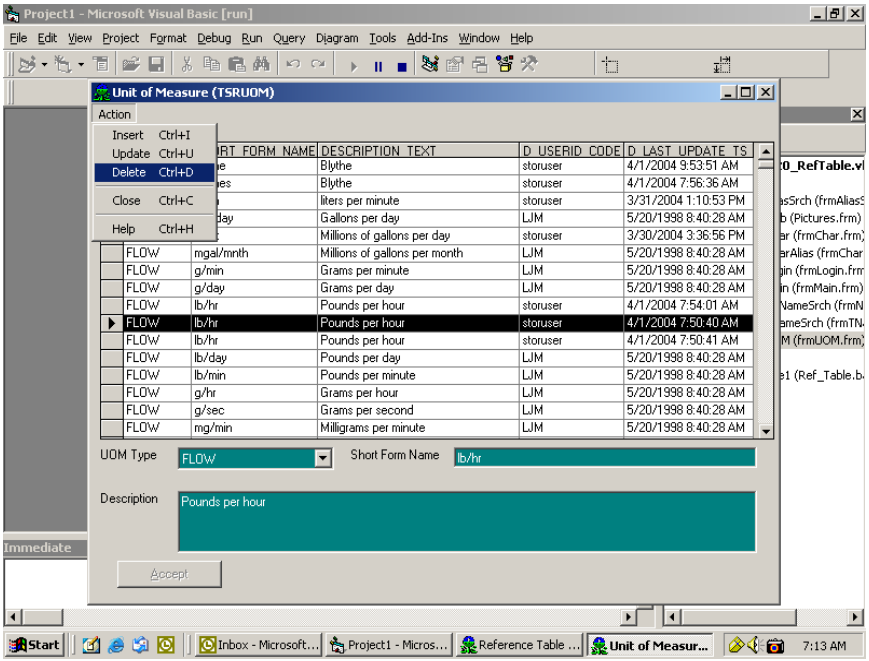
Table drop down to include all tables:



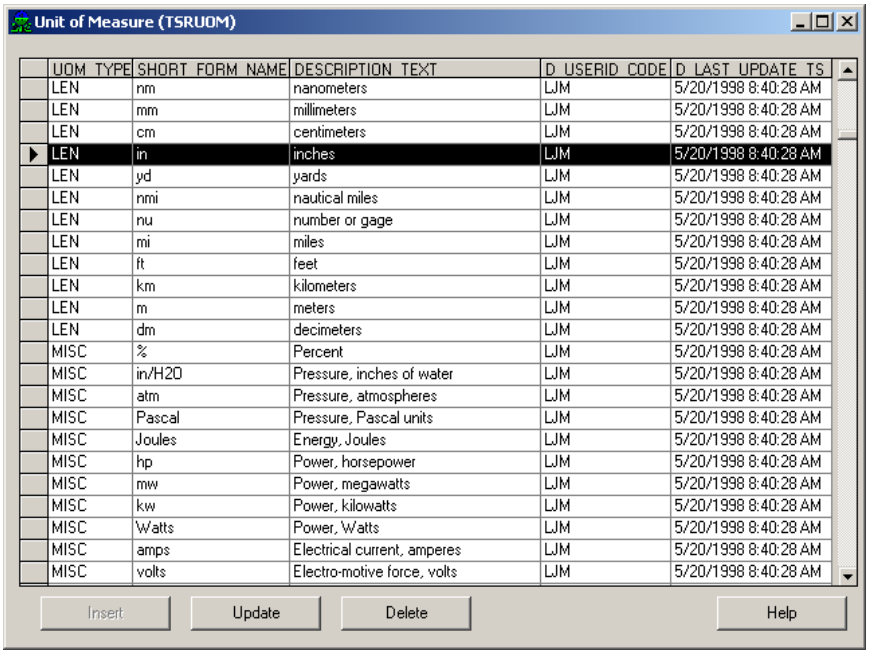
Unit of Measure Maintenance (Option 1):



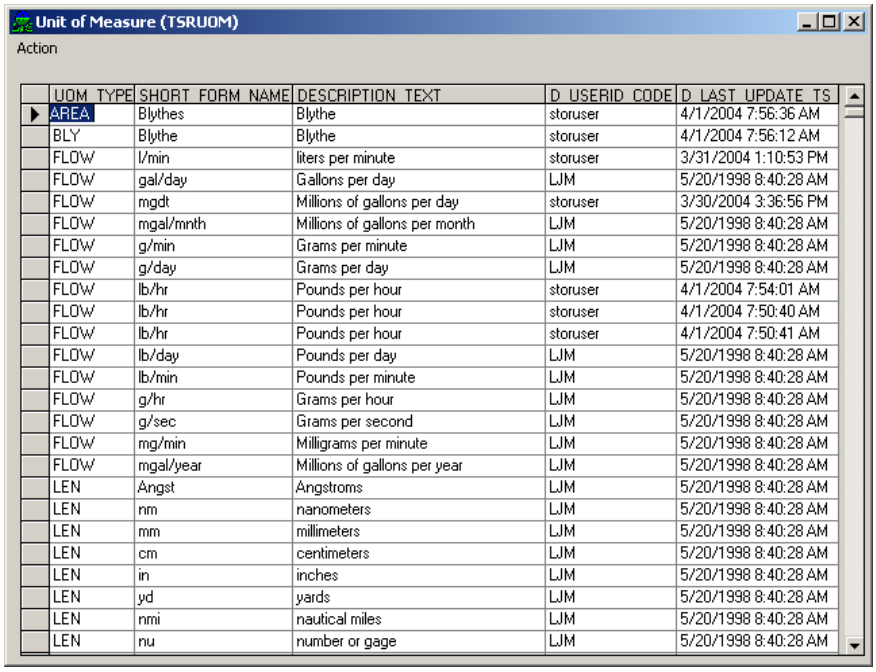
Drop down menu to perform actions:



Unit of Measure Maintenance (Option 2):

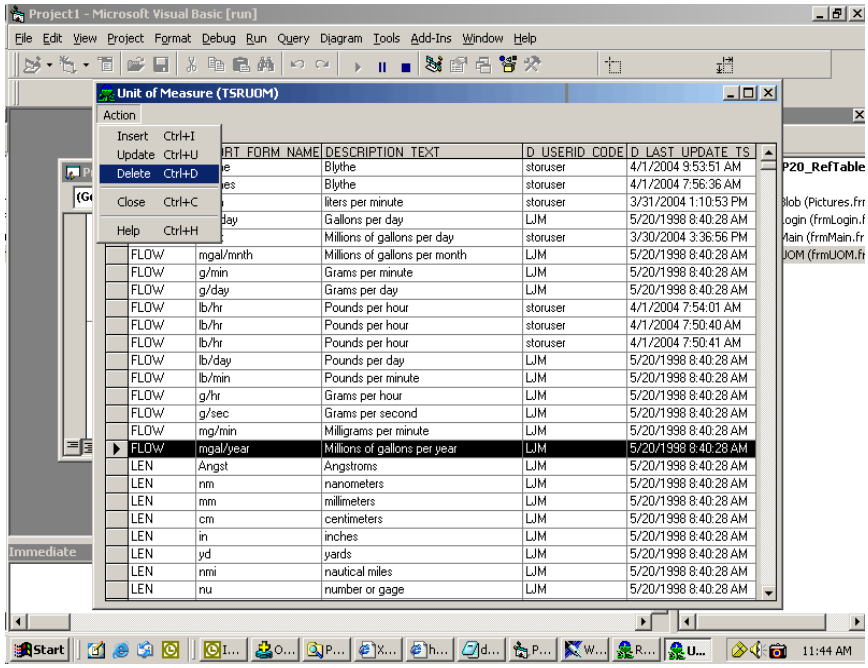


Unit of Measure Maintenance (Option 3):

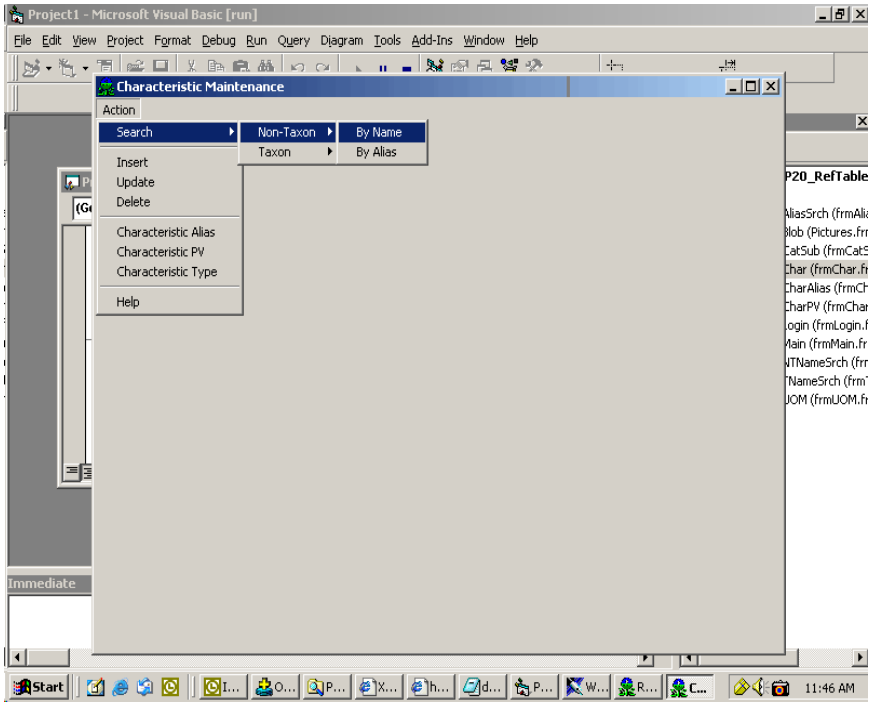


UOM_TYPE	SHORT_FORM_NAME	DESCRIPTION_TEXT	D_USERID_CODE	D_LAST_UPDATE_TS
AREA	Blythes	Blythe	storuser	4/1/2004 7:56:36 AM
BLY	Blythe	Blythe	storuser	4/1/2004 7:56:12 AM
FLOW	l/min	liters per minute	storuser	3/31/2004 1:10:53 PM
FLOW	gal/day	Gallons per day	LJM	5/20/1998 8:40:28 AM
FLOW	mgdt	Millions of gallons per day	storuser	3/30/2004 3:36:56 PM
FLOW	mgal/mnth	Millions of gallons per month	LJM	5/20/1998 8:40:28 AM
FLOW	g/min	Grams per minute	LJM	5/20/1998 8:40:28 AM
FLOW	g/day	Grams per day	LJM	5/20/1998 8:40:28 AM
FLOW	lb/hr	Pounds per hour	storuser	4/1/2004 7:54:01 AM
FLOW	lb/hr	Pounds per hour	storuser	4/1/2004 7:50:40 AM
FLOW	lb/hr	Pounds per hour	storuser	4/1/2004 7:50:41 AM
FLOW	lb/day	Pounds per day	LJM	5/20/1998 8:40:28 AM
FLOW	lb/min	Pounds per minute	LJM	5/20/1998 8:40:28 AM
FLOW	g/hr	Grams per hour	LJM	5/20/1998 8:40:28 AM
FLOW	g/sec	Grams per second	LJM	5/20/1998 8:40:28 AM
FLOW	mg/min	Milligrams per minute	LJM	5/20/1998 8:40:28 AM
FLOW	mgal/year	Millions of gallons per year	LJM	5/20/1998 8:40:28 AM
LEN	Angst	Angstroms	LJM	5/20/1998 8:40:28 AM
LEN	nm	nanometers	LJM	5/20/1998 8:40:28 AM
LEN	mm	millimeters	LJM	5/20/1998 8:40:28 AM
LEN	cm	centimeters	LJM	5/20/1998 8:40:28 AM
LEN	in	inches	LJM	5/20/1998 8:40:28 AM
LEN	yd	yards	LJM	5/20/1998 8:40:28 AM
LEN	nmi	nautical miles	LJM	5/20/1998 8:40:28 AM
LEN	nu	number or gage	LJM	5/20/1998 8:40:28 AM

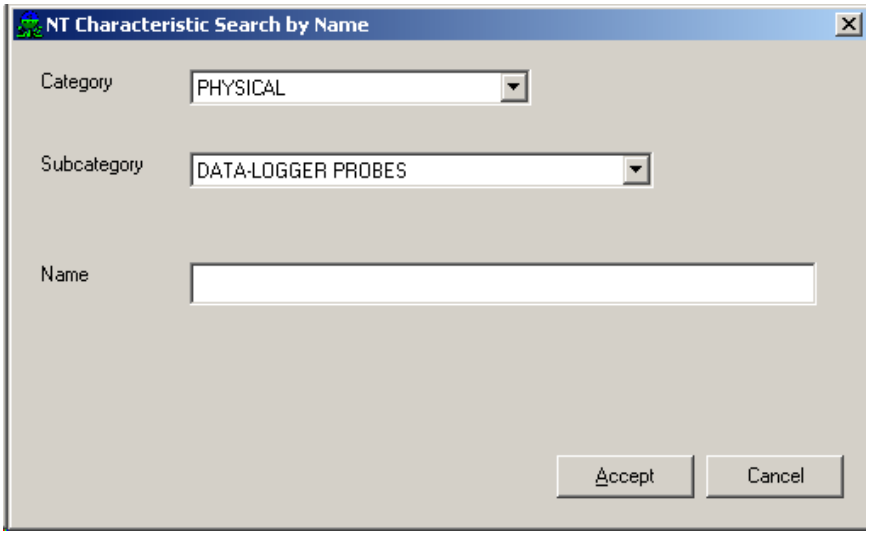
Drop down menu to perform actions:



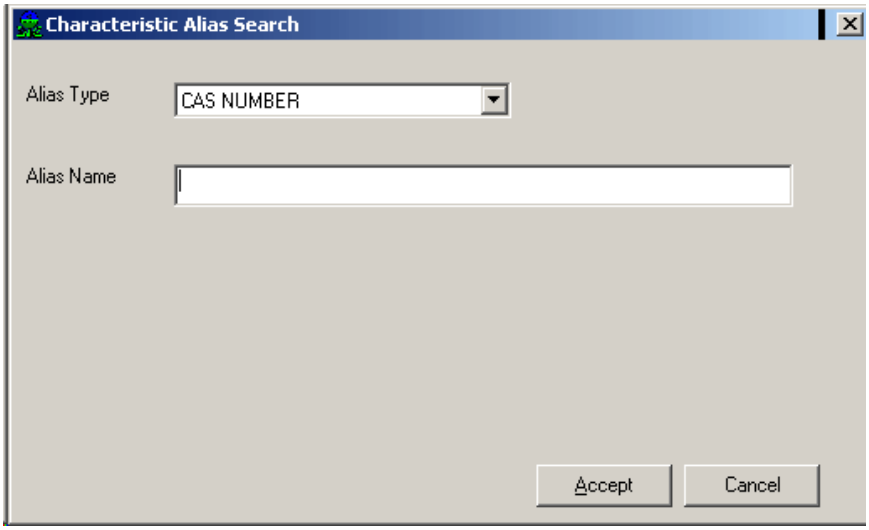
Characteristic Maintenance:



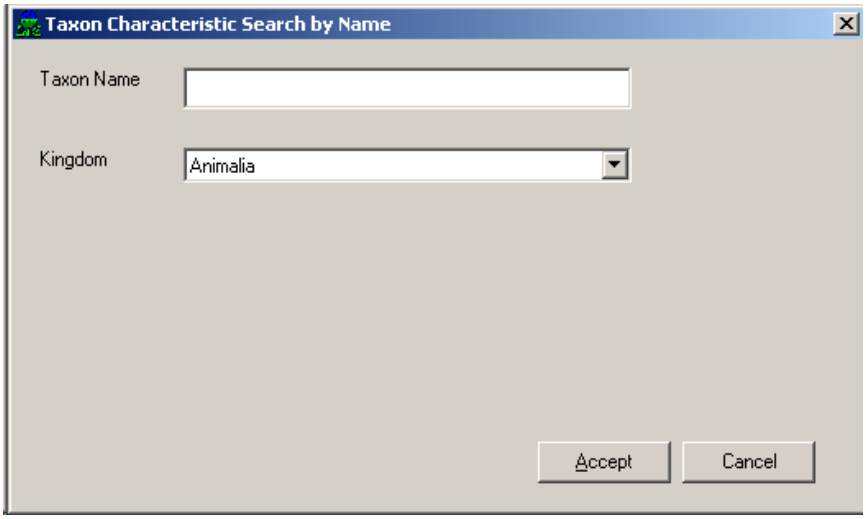
Non-Taxon Characteristic Search by Name Dialog Box:



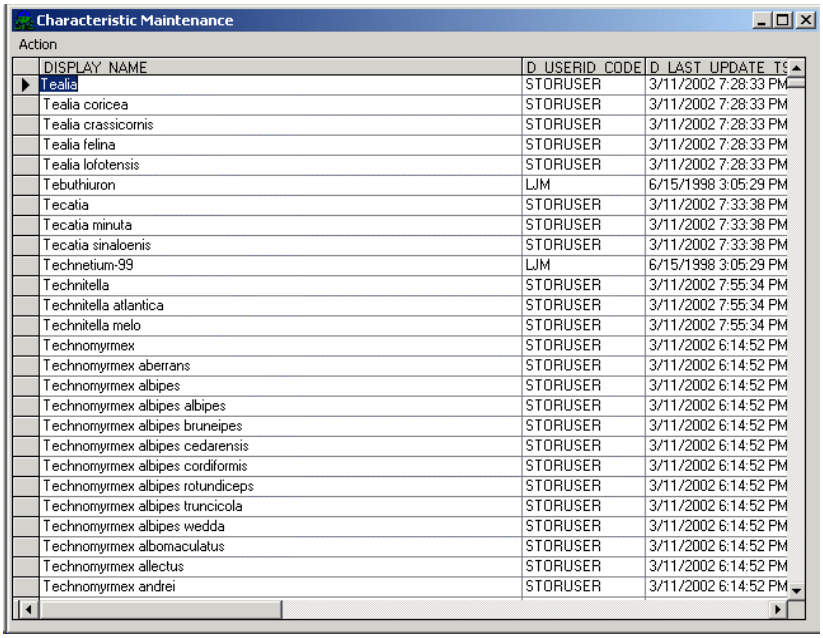
Characteristic Alias (Taxon or Non-Taxon) Search Dialog Box:



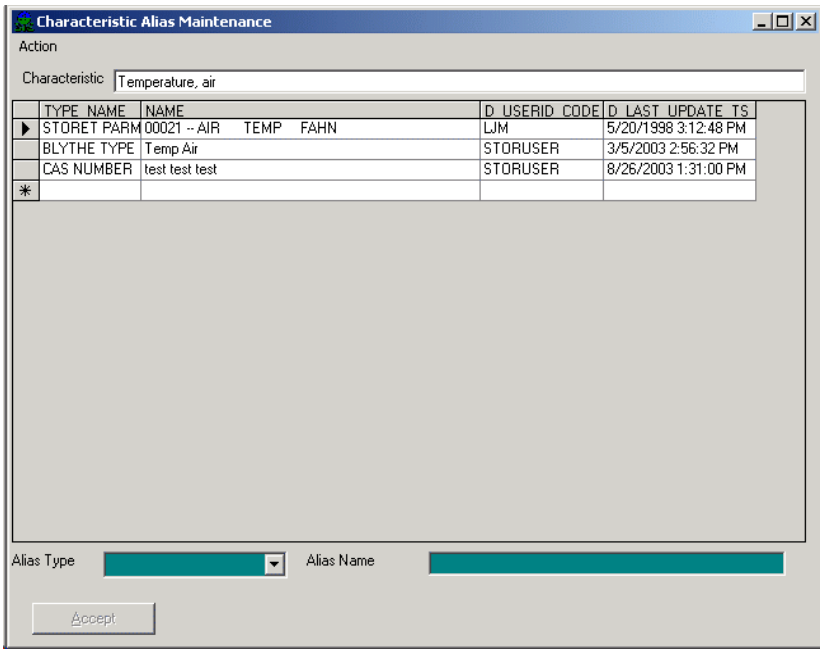
Taxon Characteristic Search by Name Dialog Box:



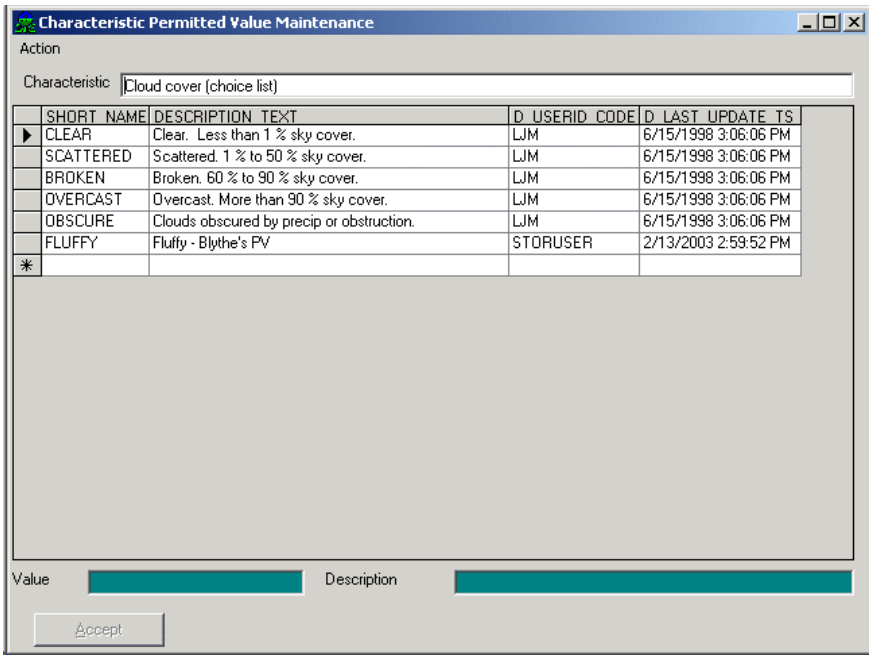
Characteristic Maintenance list now visible and populated from search:



Characteristic Alias Maintenance:



Characteristic Permitted Value Maintenance:



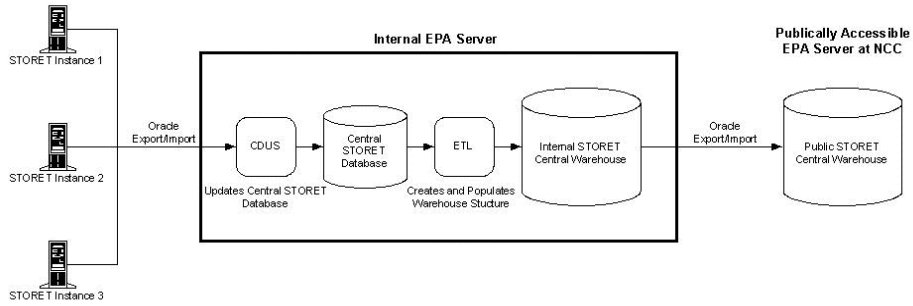
ATTACHMENT C

Transformation Options

STORET Data Transformation Options

Presented by
Joseph M. Wilson
Science Applications International Corporation

Data Acquisition
Current STORET Architecture

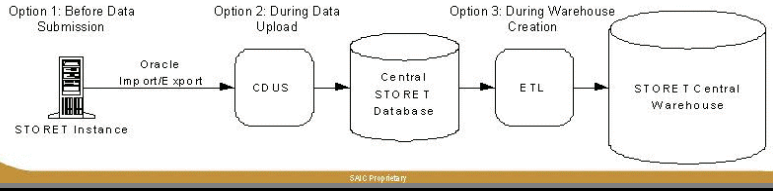


Data Transformation
Conversions and Cleanup

- Result Unit Conversions
- Datum Conversions
- Data Cleanup
  - Positive/Negative Latitudes and Longitudes
  - Duplicate Records
  - Line Breaks in Text Fields

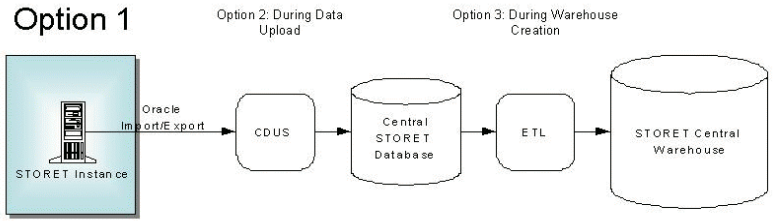
Data Transformation
Data Flow to the Warehouse

- Data Transformation and Validation Options:
  - Option 1: Before data submission
  - Option 2: During data upload
  - Option 3: During warehouse creation
- Options are not Mutually Exclusive
- Different Data Transformations Favor Different Options
- CM Control Issues must be addressed with all 3 Options



Option 1: Before Data Submission

- STORET users transform their data prior to submission

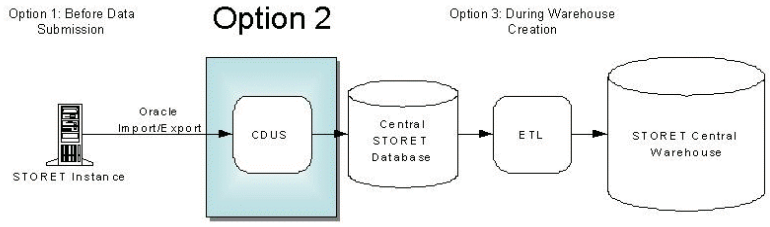


Option 1: Before Data Submission

- Advantages:
  - Effect on CDUS and ETL processing time is minimized
  - Benefits can be leveraged from all STORET applications, not just the Central Warehouse
  - Data transformations are only executed once
  - Central Database and user instances maintain a common data structure
- Disadvantages:
  - Software to perform data conversions will need to be distributed
  - Not all submitting users may download upgrades and patches
  - Data checks must still be performed later in the process to guarantee all data conversions have occurred

Option 2: During Data Upload

- Data is transformed as part of the CDUS process as new submissions are uploaded

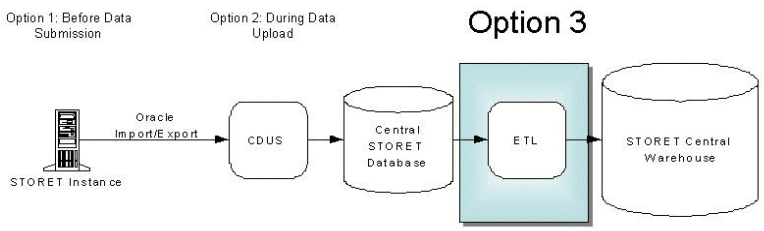


Option 2: During Data Upload

- Advantages:
  - After initial run, transformations are only performed on new data submissions
  - Eliminates the need to distribute software updates to STORET users
  - Transformed data is available from the Central Database and the Central Warehouse
- Disadvantages:
  - Requires data transformations to be performed each month on all new data submissions
  - Increases monthly CDUS processing time
  - Data transformations are not reflected on users' copies of STORET
  - Central Database and user instances have different data structures

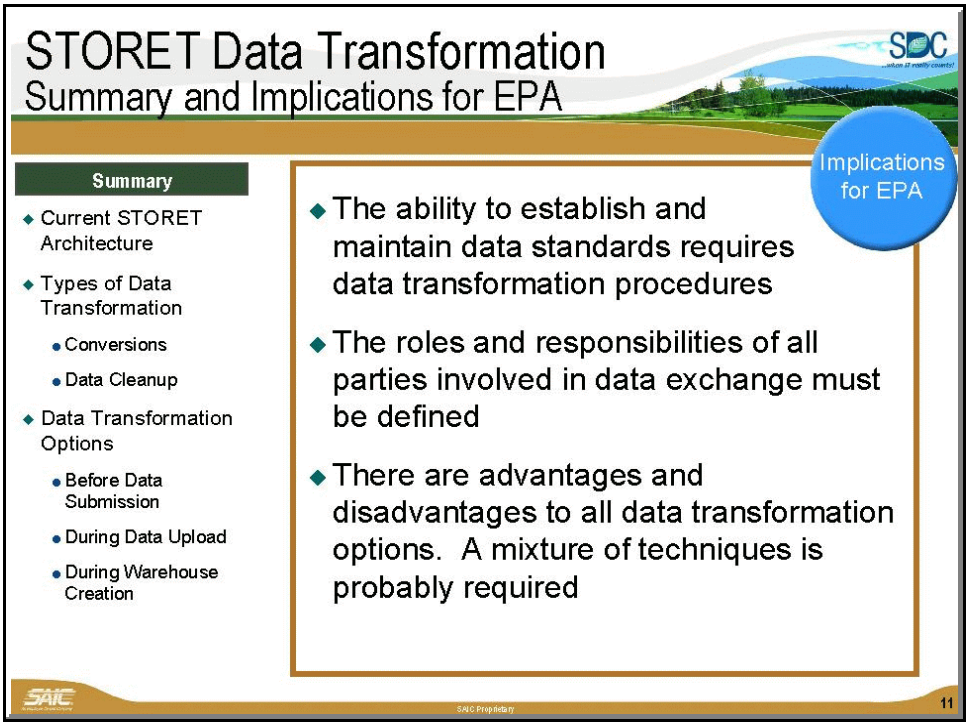
Option 3: During Warehouse Creation

- Data is transformed as part of the Data Warehouse creation process



Option 3: During Warehouse Creation

- Advantages:
  - Eliminates the need to distribute software updates to STORET users
  - Most likely the easiest to implement
  - Does not require data model changes to the STORET/STORET1 schemas
- Disadvantages:
  - Requires all data transformations to be performed each month
  - Increases monthly ETL processing time
  - Data modifications are only accessible from the Central Warehouse


STORET Data Transformation
Summary and Implications for EPA
SDC
Summary
Current STORET Architecture
Types of Data Transformation
Conversions
Data Cleanup
Data Transformation Options
Before Data Submission
During Data Upload
During Warehouse Creation
Implications for EPA
The ability to establish and maintain data standards requires data transformation procedures
The roles and responsibilities of all parties involved in data exchange must be defined
There are advantages and disadvantages to all data transformation options. A mixture of techniques is probably required
SAIC
SAIC Proprietary
11