Philadelphia Meeting: Future of STORET

Meeting Background:

A WQX regional outreach meeting was held at the EPA R3 offices in Philadelphia, PA on June 7-8, 2006. The purpose of the meeting was for EPA HQ and EPA R3 to meet with WQX stakeholders within R3 (see attendees list) for the purpose of discussing the new WQX system. Three main topics were covered:

- an overview of what the production WQX system is, and how it will function, including an overview of the WQX schema
- feedback on the data elements within the schema for purposes of sharing water quality monitoring data
- discussion of the transition to WQX for users (either from the current STORET system, or other)

The following minutes capture the issues and questions brought up during the meeting.

Day 1 June 07, 2006
1:00pm Introductions

John Armstead- Deputy Director welcomed everyone to the third of five WQX outreach sessions.

Larry Merrill went over some administrative points

Dwane Young thanked the various meeting coordinators and did introductions.

1:30pm The Future of STORET
Water Quality Exchange- Kristen Gunthardt gave presentation giving an overview of WQX
- EPA's Vision for WQX
- WQX Context
- OWWQX Pilot
- WQX Schema
- Schedule for Implementation

Existing Tool Implications: Questions/Answers
Q-Will data already in STORET be migrated to WQX?
A-You can load data yourself or you can work with EPA to move data to warehouse.
    Q-What about Legacy Data?
    A-Data can still be retrieved as long as it continues to work.
Q-How many states have a fully functioning node?

A-Region 3 exchange network person will address tomorrow.
        Q-What if we don't have a node?
        A-Smaller groups will use web services.
Q-How did the tribe do it?
A-Generated XML, Mapped to the schema, sent through CDX integration test tool,
        Q-Authentication handled by Individual Organization or by Node?
        A-Authentication will probably be handled through CDX.
Q-Where does Beaches fit?
A-WQX team will meeting with them next week. No reason it cannot be worked out.
        Q-Is ESAR a multimedia standard (are other agencies using it)?
        A-Superfund, Emergency Response and others are moving toward ESAR.
Q-What is the life cycle of this new version?
A-Technology is difficult to keep up with because applications quickly become obsolete.
WQX is instead moving toward a standard that does not change.
        Q-I want to hear Kevin say old flow will still work until 09?
        A-"No problem."
Q-With requirements to put data into STORET, is transition a worthwhile endeavor?
A-It depends on the schedule of your grant. Any effort made to put data into STORET is
not effort lost, since WQX is similar.
        Q-Will biological include bio-toxicity?
        A-Perhaps.
Q-Where are other federal agencies with this, NOAA, NWIS, USGS?
A-USGS- discussion to map NWIS with STORET (currently dished out side by side in
window to my environment) hopefully to be able to publish both using same schema
NOAA- a little further on the radar screen- if you have any inroads please give contacts
        Q-How many States serve up data via their own system?
        A-6-7, some are real time.
Q-If you shift to Enviromapper away from Window to My Envt, how is this integration?
A-WME is still there but we are not investing any further updates at this time.

3:00pm Break

3:15pm Introduction to the WQX Schema

**WQX Schema Structure- Dwane Young went over the WQX schema structure and
data elements. (schema current as of June 2, 2006)**

Schema Structure & Data Elements: Questions/Answers
Q-Do you prefer to keep your old ORG IDs?
A-Yes! You can use old names but some punctuation marks may become illegal.
        Q-Reaches- Segments can be indexed with NHD- Make this optional?
        A-YES YES YES
Q-Does WQX allow activites to be tied to multiple projects? A-Yes
        Q-Can we keep the same activity ids?
        A-Maybe- need to be unique across entire org
Q-Do I have to make a vertical measurement each time?

A- No, now it is optional.

Q-Estuaries-still primary/secondary- A- no, they longer exist

Q-Can Projection be a new data element? A-We will put it on the issues list.

Q-Can Photo scale/horizontal accuracy be a new data element?

A- We'll put on issues list.

Q-What about Time Zone?

A-W3C standard for XML is all times based on Greenwich/UTC with an offset- come up with list of counties/timezones to calculate on the fly or just create a new required data element called timezone.

Q-Can Result Depth be a new data element at result level?

A-We'll put it on the issues list.

Q-Why are we having org name instead of org id in activityconductionorgtext?

A-Only for if organization different from normal org and is text.

Q-Tissue sample only taxa name and organ name?

A-Others will be added in biological

Q-Nondetects now do not strip units- should units remain? A-YES

Q-Add actual value of nondetect as new data element?

A-Will put on issues list

Q-Sample fraction- want it to remain mandatory? A-YES Metals analysis needs it

Q-05000501- How do you handle weather if size is too small? A-??????

Q-Analytical Methods are conditional on having a non text result-req?

A-YES but with no controls is there value here?

Q-Need lab certified as a new data element? A-Perhaps not

Procedures for incorporating WQX Schema feedback – Dwane Young went over these procedures:

Background Questions:
- Is it necessary for sharing data?
- Is it part of the ESAR standard?

Biological Data
- WQX is waiting for biological data standards (Question about whether WQX will take unilateral position or wait until they are done? We will give Standards groups first crack at developing them)

Versioning problem from Schema changes- consider having schema date ranges

Suggest creating a user group committee to help bounce ideas regarding changes.

Day 1 Closing thought:

Required/Optional/Conditional Fields

Data Flow- When you submit a file to the schema, you also have to provide all the elements that define org and project (location id, lat/long, etc) even if it already exists in WQX. Do you want to be able to submit chunks without providing all the other data? Validation will occur at back end. Think about tonight and will re address tomorrow.

4:30pm Adjourn

Day 2

9:00am   WQX Schema- Dwane Young
Some concerns were brought up from Day 1:
- Concerned about data being handed over and forgotten (Trash Can)- need ability to update results so corrections are easier- XML allows to flag changes since last submission- changes to core ids could cause duplicates-
  - Does WQX tag flags as updates? No, if different just assumes it is an update
- How to avoid duplication across systems? USGS- same location?, time? Samples?
- Future concept of pulling data from multiple systems.
- Possible solution to flag data as existing in other systems.

Domain Values- Kristen Gunthardt went over the domain values lists.
For a complete set of domain values, go online to the Future of STORET site.

WQX Data Flow- Ryan Jorgensen from Gold Systems
Ryan went over in greater detail the submission process and data flow for WQX. This includes presenting information about a Node, CDX web site or XML Generation Tool WQX submission and the associated web services.

Does system support transaction load- rollback? Give me history on my submissions transaction log (maybe)

Overview of XML- Dwane Young gave a short overview of XML, and how we use it to submit Water Quality data.
10:30pm Break

Schedule/Data Flow

Dwane went over the data flow again in the framework of our production schedule:

Dwane also highlighted the following web services that we are planning to provide from the backend WQX database:

1. Get project- query CDX to return projects loaded
2. Get Monitoring Location- query CDX to get Monitoring Locations by Your org or Project
3. Get Activities- # Results, # Preliminary, (date as another parameter?)
4. Get Results- ask for activity and get results (date range? Blobs in queue?)
5. Get transactions- give me all transactions from date range (add transaction id?)
6. Get domain values

The following are web services from warehouse by March 2007
1. Get stations
2. Get results

3. Get activities as a potential

Browser interface for non node users? Enviromapper update
Spatial querys?
NOTE: Air had performance related issues when entire datasets asked for

If they can generate their own XML but have no node, can they use CDX node? perhaps
Can community node be created for multiple users? Yes – pacific north west example

11:00am Local Issues and Data Management Requirements

So what does this mean to a user?

Identify challenges for transition

Data Management alternatives

12:00pm Lunch

1:00pm   Local Issues
Schedule- STORET 2009 is a workable deadline for most

During this session, the following challenges were listed for the different user types:

B/C Users (Those users not using STORET as their primary data management system)
Mapping to the Schema
          Available in Access
          Map to ODE
Timeframe- Release final Schema earlier (moving target)
BLOB issue- Class C needs a way to accommodate BLOBs
Authentication for Multiple Organizations
          Ship multiple orgs from same node? Yes
          Can smaller groups prevent data from being seen? Maybe
          Data from multiple orgs in same XML steam? No
User Participation- We want a vote!!!
Legacy Data
          Want to put into WQX
          Continued Outreach
Simple preferred/flexibility- value of WQX is its simplicity- concern over time that it will become too complex with burdensome required fields
QA
          Range Checks- future wishlist
          Accommodated in overall process, different levels,
$$$
          Burden shifting

Seed Money- to support development of own system

A Users (Those that use STORET as their primary data management system)

What do we do now without STORET as a data system?
    Options
Communication
    Program Staff/IT Staff/Monitoring Staff
Good Ideas
    What can we learn from Air?
    How to coordinate with Regional IRM(InfoResourceManager) Chief
    WQX is Equivalent to STORET for grant purposes.
    Monthly conference calls to facilitate continued discussion
    Region 8- regionwide system in the works
    Consortium- system to meet groups needs (most here have local needs so not
as appropriate)
    State approaches up on the web to share
    EPA keeps changing their minds- "its gonna be there forever"

Numerate alternatives / Warehouse suggestions
    Query national database- give me all PCBs (already done through warehouse)
    SAS Proc with statistical analysis tools
    Consensus on low level statistics about raw data
      (precalculated- counts, means, max/min)
    Results on screen
    Canning old inventory
    Show me stations that exceed certain values
    List of parameters/variables that are indexed- top end
    What types of questions are being asked?
    How are the data used?—describe in public executive summary form—arcims
    Map interface is critical
    Anonymity- people prefer not to have prompts
    Enhancing the data out flow
    Parameter frequencies- are we tracking them?
    Download data from bay- need to fill out a profile
    Problems having basic statistics on methods- comparability difficulties

2:00pm   EPA Assistance and Developing a Transition Plan
EPA's role in the transition

NEIEN Grants-Joe Kuntz- Network Grant Coordinator for Region 3
Works with HQ to review/write guidance- panelist to decide which grantees get grants
4 categories
    Implementation- generally 225K-1M
    Challenge- cap 1M
    Mentoring

Administrative- mostly goes to ECOS

Who in your state is the main state contact

ExchangeNetwork.net is a complete state profile with operating nodes –
40 out of 50 with contact names as well
Also by data system
How to build your own node-

Epa.gov/OEI navigate to state/grant--Guidance for grants and applications come out Mid
November, awards usually occur July August, review Jan Feb 2007
- unsure what budget will be- OMB pushback due to unliquidated funds 2002

Funding support to establish nodes and pay for data flows- thirdly foster novel approach

Internally- individual programs contacting main state are more successful/coordinated

Do challenge grants allow for orgs outside the US? In principle they can, lead state
agency takes and distributes the funds

Budget of 18-20M a year

Toolbox for transition assistance- Kristen

Explain relationship between WQX and CDX- CDX in OEI but WQX in OW

5 Options

XML Generation Tool- non node user method to generate XML
CDX integration test tool- used during pilot process for smaller users
Nationally Hosted WebSIM- support tribes until XML generation tool up
Transition support funds through OW- Seed Money
NEIEN Grants through OEI
Infrastructure
Data Exchange
Mentoring
Challenge Grant

Elements of transition

Schedule- SO WHAT?

1 determine whether or not you have a node
2 if you have a node, is part of that grant that you will submit WQ data?
3 coordinate with IT staff for schedule of when to flow data to WQX

1 pick an option for replacing storet
2 what are your reqirements
3 where will you get the resources (as EPA)

15 heavy storet state users so money put aside to help transition

1 begin development of new system within next 2 years
2 continue to compare your system against WQX (don't wait for biological)
3 coordinate with EPA on data migration

1 begin mapping your data to WQX schema
2 work with IT staff to set up data system to send data via the node
3 send us data

MOST IMPORTANT THINGS TO HELP US TRANSITION
Test environment for WQX- just list everything as preliminary
Constraint on file size? 30megabytes used in pilot but may be expanded
Put all presentations on website
Document schema changes on website- storet website and also on exchange
network doc including flow configuration doc, etc
Create list of Water quality systems that are available- states who have already
looked- use list serv
VA, MA, DE has great system they will share
Migrate the data from the current Modern STORET system
STORET can still be used

3:00pm Wrap Up

Discuss outcomes
Keep Running Storet- cost to generate XML- will continue to work- Users
responsible for keeping domain values up to date
ODS- give model as a starting point
Other states have system that you like
Timeframe- Schema releases 95% completed, til outreach completed
Node – XML generation tool- participation in design
Communication- how can EPA help with communication- peter talk to
counterparts who work 303d, 305b, ADB,
Emphasis to get other EPA programs to put data into warehouse- Chesapeake
bay, Great Lakes, EMPAP data, estuary systems
XML tool from storet structure
States will be looking for guidance
Domain Lists- method is just free text- how many submitters are also users- if
you make it too easy/relaxed there is a danger
Timing issue for WQX flow- if you make the XML generation tool to read a
flatfile and generate XML, it is universal and sharable- if you are reading a
database, then they are not platform independent
If you have a node, then you should flow data through the node.

4:00pm Adjourn

Action Items

1. Add relative depth data element (EPA)
2. Reconsider the source map scale data element rule – should this be required for
   other methods of determining location? (EPA)
3. Research adding Depth at the result level; PDL depth as well(EPA)
4. Revisit the Result analytical method business rule – should this element be
   required?

Attendance

Dwane Young	EPA HQ	Kristin Gunthardt	EPA HQ
Peter Grevatt	EPA HQ	Rome Arquines	EPA HQ
Ryan Jorgensen	Gold Systems	Ed Kim	EPA Region 1
Alexander Yeloah	MDE	Paul Morton	NJDEP
Mike Kusmiesz	NJDEP	Harry Chen	NJDEP
Megan Tulloch	RTI	*Mike Beaushene	CTDEP
*Mellony Hoskinson	RTI	Grace Hanrahan	EPA Region 3
Deb Forman	EPA Region 3	John Graves	EPA Region 3
John Armstead	Philadelphia(Deputy Director)	Bryan Carr	WVDEP
Carlton Washburn	MDDEP	Phil Duyinski	PWD
Elinor Zetina	MDMDE	*Dawn Hintz	SRBC
*Brydon Lidle	SRBC	Brian Chalfont	PADEP
*Alex Trounov	Tetra Tech	*Tony Allred	MDDNR
*Mike Townshend	DEDNREC	*Kathy Knowles	DEDNREC
*Deb Soule	NHDES	*Peter Rushton	MEDEP
*Kimberly Cornelison	LADEQ	*Chris O'Connor	NYDEC
*Mike Long	EPA Region 3	*Brian Burch	EPA CBP
*Donna Beatsch	ORANSCO	*John C. Klear	ORANSCO
Degeuz Steuart	VADEQ	Vimala John	MDE
*Lee Manning	Potomac Appalachian Trail Club	Stan Stepanson	EPA Region 2
*Nancy Ragland	TCEQ	Ernestine Bryant	LMIT
Kevin Christian	EPA HQ	Hector Sqariabo	Virgin Islands
*Nicholas Protospaltis	Interstate Env. Comm	*Jefffrey White	Tetra Tech
*Roger Stewart	VADEQ	*Steve Tessler	USGS

*Indicates a B/C user