Summary Report of
HydroXC Phase 3 Activities

NOAA: NWS – Office of Hydrologic Development

A summary of activities conducted between November 2006 and August 2007

Stephanie Liu-Barnes, Director of Consulting Services
Glen Oliff, Senior Application Engineer
Alexis Karnauskas, HydroXC Coordinator

August 31, 2007

Table of Contents

Background

The National Weather Service's Office of Hydrologic Development (OHD) and Apex Digital Systems (Apex) have worked together for over two years on establishing and growing the Hydrologic XML Consortium (HydroXC). The purpose of the consortium is to create a common, self-documenting method for exchanging hydrologic data among the many organizations that use such data both operationally and for research.

In November of 2006, OHD tasked Apex with a third phase of activities to begin to put into place more practical advancements on the theories and general schemas derived in earlier phases. Feedback we heard from Consortium members during the second phase asked for pragmatic requirements for using the XML schema, and for making it more specific to daily needs of hydrologic software users.

Engagement Overview

Phase 3 of the HydroXC work was divided to address three main areas: research, development, and community building. The tasks required for each area are as follows:

- **Research**: The research component of this phase was intended to further evolve the HydroXC XML schema. To do so, the HydroXC team created a few basic hydrologic object representations. Using these new structures, and with further discoveries from the data adapter work, we updated version 2.0 of the HydroXC schema to version 3.0.
- **Development**: The development component of this phase was intended to create a proof-of-concept data adapter capable of reading and writing messages between the SHEF unpacked .B format and HydroXC-compliant XML. This component directly addresses the need for developing basic tools and examples for hydrologic organizations to begin using for their own modeling and data delivery. The hope is that this example data adapter will provide guidance for adapter creation for other groups interested in developing adapters for sharing data in other formats.
- **Community Building**: The community building component of this phase was intended to strengthen the collaboration among the HydroXC members. Activities included upgrading the HydroXC website, including Michael Piasecki, PhD., in core planning and work, presenting HydroXC at a conference, conducting two web workshops, and sending monthly newsletters alerting the members of new progress.

This document summarizes the activities and accomplishments from each of the three areas and describes suggested and appropriate next steps for HydroXC.

Research

The intent of the research activities was to further evolve the HydroXC XML schema. This began with the creation of several hydrologic object representations and was further informed by the data adapter proof of concept work.

OHD and Apex brought in a partner to assist with this work – Michael Piasecki, PhD, an Associate Professor of Civil & Architectural Engineering at Drexel University who specializes in hydroinformatics and water resources modeling. Dr. Piasecki had participated in earlier phases as a general member of the consortium, but he and his team of graduate students played an instrumental role in this phase by providing technical expertise and thought leadership for all HydroXC discussions.

Dr. Piasecki's research interests include the development of Information Systems in Hydraulic and Hydrologic Engineering and Environmental Decision Support Systems using web-based technology and the development and application of numerical models for free-surface flow domains in open channel flow, surface water hydrology, and estuarine environments. He is also involved with CUAHSI, the Consortium of Universities for the Advancement of Hydrologic Science. CUAHSI is an organization representing more than one hundred U.S. universities with a mission to foster advancements in the hydrologic sciences. Dr. Piasecki and the graduate students in his lab support the Hydrologic Information Systems (HIS) effort at CUAHSI.

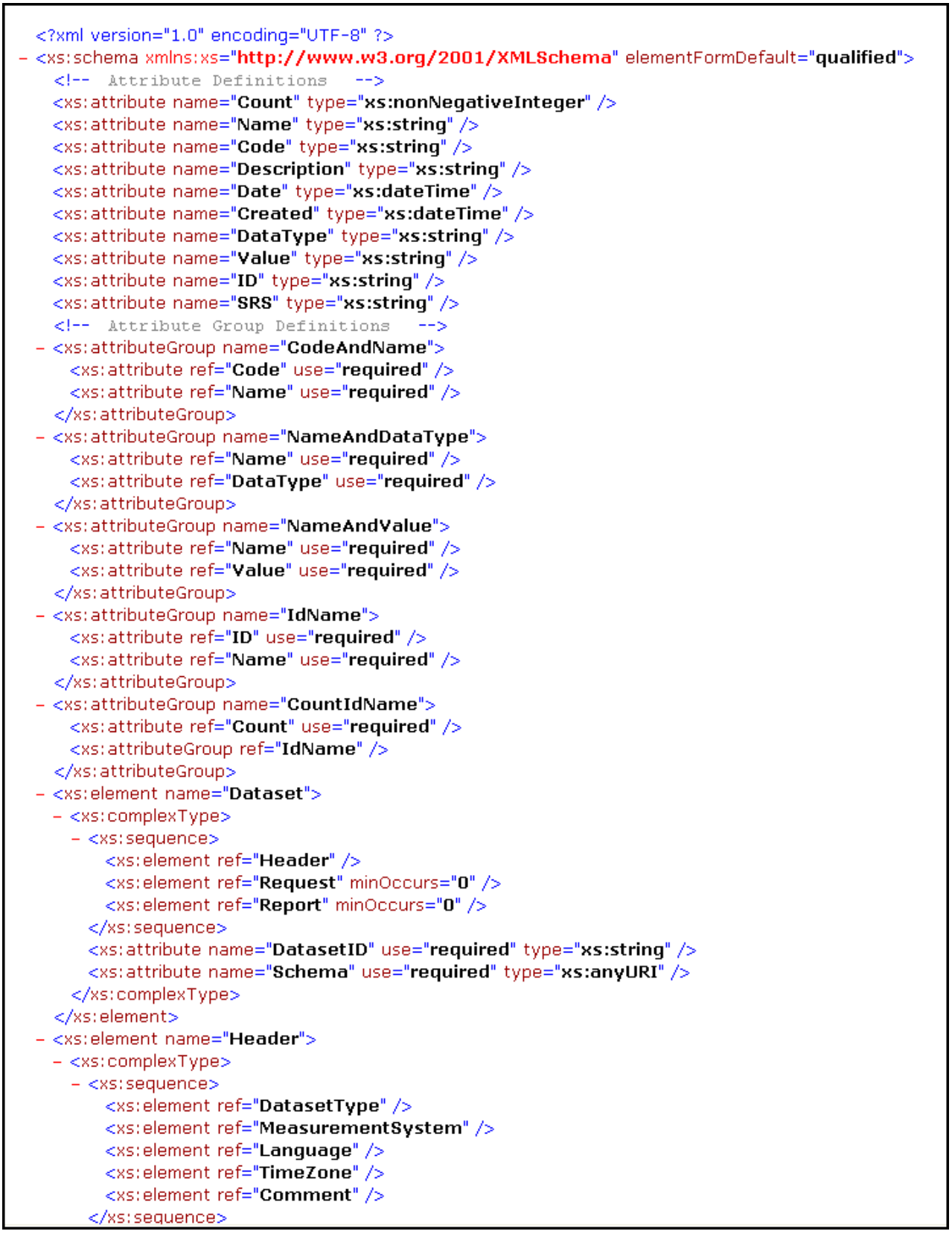
Schema Version 3.0

The HydroXC schema that resulted from the Phase 2 efforts was simply a block diagram representation documented using Microsoft Visio. In order to fully create a working HydroXC schema and make it usable for operational and research-oriented hydrologic groups, Dr. Piasecki's group transformed the basic version 2.0 schema into a fully qualified XML schema file (.xsd). Not only did the schema improve in its documentation, but the overall structures also evolved to accommodate for the additional features realized during the creation of the hydrologic object representations. The updates included improved location data handling and renaming of some of the supporting structures for better specificity. A screenshot of the version 3.0 schema (.xsd file) is shown below for reference. This file is posted to the HydroXC website, http://www.hydroxc.org, for overall dissemination and access.

Hydrologic Object Representations

With a format as flexible as XML, we determined there needed to be suggested structures for documenting certain basic hydrologic objects in the HydroXC XML format. In other words, clear benefit could come from all members documenting a reservoir, for example, in the same way. While these objects will likely continue to evolve over time, Dr. Piasecki and his team created basic representations for the following hydrologic objects:

- Channel cross section
- Rating curve
- Reservoir
- River reach



```
<?xml version="1.0" encoding="UTF-8" ?>
- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
    <!--  Attribute Definitions  -->
    <xs:attribute name="Count" type="xs:nonNegativeInteger" />
    <xs:attribute name="Name" type="xs:string" />
    <xs:attribute name="Code" type="xs:string" />
    <xs:attribute name="Description" type="xs:string" />
    <xs:attribute name="Date" type="xs:dateTime" />
    <xs:attribute name="Created" type="xs:dateTime" />
    <xs:attribute name="DataType" type="xs:string" />
    <xs:attribute name="Value" type="xs:string" />
    <xs:attribute name="ID" type="xs:string" />
    <xs:attribute name="SRS" type="xs:string" />
    <!--  Attribute Group Definitions  -->
  - <xs:attributeGroup name="CodeAndName">
      <xs:attribute ref="Code" use="required" />
      <xs:attribute ref="Name" use="required" />
    </xs:attributeGroup>
  - <xs:attributeGroup name="NameAndDataType">
      <xs:attribute ref="Name" use="required" />
      <xs:attribute ref="DataType" use="required" />
    </xs:attributeGroup>
  - <xs:attributeGroup name="NameAndValue">
      <xs:attribute ref="Name" use="required" />
      <xs:attribute ref="Value" use="required" />
    </xs:attributeGroup>
  - <xs:attributeGroup name="IdName">
      <xs:attribute ref="ID" use="required" />
      <xs:attribute ref="Name" use="required" />
    </xs:attributeGroup>
  - <xs:attributeGroup name="CountIdName">
      <xs:attribute ref="Count" use="required" />
      <xs:attributeGroup ref="IdName" />
    </xs:attributeGroup>
  - <xs:element name="Dataset">
    - <xs:complexType>
      - <xs:sequence>
          <xs:element ref="Header" />
          <xs:element ref="Request" minOccurs="0" />
          <xs:element ref="Report" minOccurs="0" />
        </xs:sequence>
        <xs:attribute name="DatasetID" use="required" type="xs:string" />
        <xs:attribute name="Schema" use="required" type="xs:anyURI" />
      </xs:complexType>
    </xs:element>
  - <xs:element name="Header">
    - <xs:complexType>
      - <xs:sequence>
          <xs:element ref="DatasetType" />
          <xs:element ref="MeasurementSystem" />
          <xs:element ref="Language" />
          <xs:element ref="TimeZone" />
          <xs:element ref="Comment" />
        </xs:sequence>
```

Figure 1: Snapshot from the HydroXC Schema Version 3.0, published May 29, 2007

Development

The development component of the Phase 3 work involved creating a proof-of-concept data adapter capable of reading and writing messages between the SHEF format and HydroXC-compliant XML. To enable hydrologic organizations to begin using the schema for their own modeling and data delivery, members are in agreement that it is essential to begin developing these basic tools and examples.

The hydrologic community is attempting to move towards a higher degree of interoperability by trying to find a mechanism to better exchange hydrologic data. Currently every organization and group tends to have its own format, which requires parties wishing to share data to be able to read and understand each individual proprietary format. For example, if a non-NOAA hydrology professional wishes to use SHEF data, they must learn the format and build custom tools to process it, or they must convince a NOAA user to translate the data for them. Likewise, a similar process must occur if NOAA wishes to use another organization's data in their proprietary format. This involves a time-consuming "ramp up" process and hinders the ability for agencies to share information, therefore slowing the overall advancement in hydro-meteorological science.

The purpose of this proof of concept was to demonstrate how, using HydroXC XML as the standard format, data adapters could be built to translate data to and from a proprietary format and HydroXC XML. In an effort to keep the scope manageable for short-term progress, the proof of concept was limited to one type of SHEF message – the unpacked .B format.

Proof of Concept SHEF-HydroXC Data Adapter

In general, the proof of concept data adapter is capable of transforming data from an existing unpacked SHEF .B format into a HydroXC 3.0 compliant XML format. Additionally, the adapter can transform data from a HydroXC 3.0 compliant XML format containing the relevant data into a SHEF .B format. While this version of the adapter is meant to demonstrate the basic concept, it also provides example code for further adapter development. For example, if someone with an unpacked SHEF .B format file wishes to share data with a user who normally deals with HEC files, they could use a SHEF data adapter to transform the data from SHEF to HydroXC, and then use a future HEC data adapter to transform the HydroXC file into a HEC file.

The SHEF – HydroXC data adapter developed is intended to provide a proof-of-concept tool for the following high-level functions:

- Enable a user outside of the NOAA environment to process and understand hydrologic data currently stored in the SHEF .B unpacked format (SHEF -> HydroXC)
- Enable a user inside the NOAA environment to process data from an outside source and format (represented in HydroXC-compliant XML), within SHEF-handling systems (HydroXC -> SHEF)

The proof of concept data adapter also utilizes the existing SHEF parser. The adapter takes as input, the SHEF parser's binary output. When transforming any new SHEF .B messages (i.e. one that is not provided already), a user must first parse the message using the SHEF parser, and then feed the resulting data stream into the proof-of-concept adapter.

The following diagram shows the basic process flow for data transformation to and from SHEF and HydroXC.

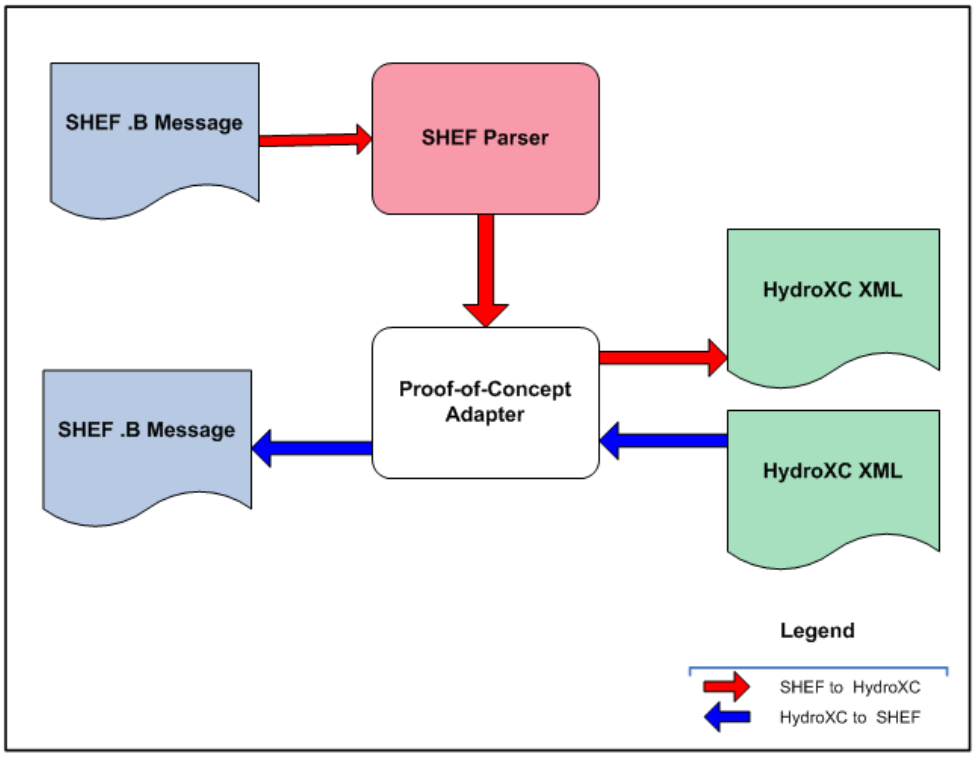


Figure 2: Proof of concept SHEF – HydroXC data adapter process

The proof of concept data adapter was created, tested, and presented to OHD, Dr. Piasecki, and the members of the consortium. While the function of the proof of concept remains basic, the Apex team was able to include a couple of configurations that will help the data adapter to be expanded at a later time. These are:

- **Location Lookup**: Using a simple XML file as an example, the location lookup file demonstrates the ability to interpret data proprietary to one data source into a format that is usable by anyone in the hydrologic community. For example, NOAA uses their own station identifiers in their files that represent a specific latitude and longitude location. To non-NOAA users, the station identifiers do not easily indicate what the corresponding lat/long values are. The location lookup file includes station identifiers with their lat/long pair and, when included in the runtime parameters, the data adapter writes the latitude and longitude values into the resulting HydroXC file.
- **Parameter Mapping**: The parameter mapping file is also an XML file that is required for transformation from HydroXC-compliant XML to the proprietary format. The purpose of parameter mapping is to enable users to transform data from an external proprietary format into the format that meets their internal needs. Currently, the SHEF proof of concept adapter handles mapping of explicit values. There is also handling for invocation of a custom algorithm for conversion of values. For example, if a data producer saves temperature readings in degrees Celsius, but a data consumer needs temperature data in degrees Fahrenheit for their systems, the parameter mapping file can convert the temperature values using a provided algorithm. This saves the users the extra step of converting the values at a later step in the process.

The resulting proof of concept data adapter is fully functioning for basic conversion needs. In these basic cases, the adapter can be used for translating data between HydroXC and unpacked .B SHEF, but should be tested thoroughly before launching into any production-level processing. Now that the data adapter has been developed, Consortium members can include the adapter in an existing application as a further proof of concept or extend the mapping capabilities so that the adapter can handle more sophisticated data situations. Additionally, members can use the proof of concept as an example to create data adapters for other proprietary formats so as to enable further use and adoption of the HydroXC format.

Forward-Looking Development Opportunities

With the HydroXC-SHEF proof of concept adapter in place, the hydrologic community can now take several next steps for advancing the functionality, usability and adoption of this technology. Ideas discussed during Consortium workshops and meetings include:

- **Invocation Framework** – Once several adapters have been built, it will be meaningful to create an invocation framework that can be easily configured for custom transformations. This framework would serve as a HydroXC transformation workflow management framework and would enable users to more easily select adapters that are necessary for their data conversion and possibly run the adapters without having to set up their own development environment. For example, we envision a framework with a web-based user interface from which a member of the hydrologic community can specify the proprietary data format, such as SHEF, and the direction of transformation, such as “SHEF to HydroXC” or “HydroXC to SHEF”. The framework would automatically select the appropriate adapter and execute the process given the input files from the user.

- **Standard Algorithm Conversion Library** – The proof of concept adapter demonstrates the value of invoking custom algorithms for connecting disparate organizations’ data formats. It would be extremely useful for the Consortium to agree on a standard set of algorithms to be used for the more common format conversions. For example, there may be standard temperature conversion and measurement conversion algorithms. These may not need to be created from scratch, but rather “blessed” from a set of existing conversion utilities and packaged along with the existing HydroXC software.

- **Refactoring of Adapters** – Once the current concepts have been exercised by implementing a few additional data adapters, it will be necessary for the Consortium to begin to identify and encapsulate any common functionality. This will enable appropriate reuse of basic functionality, thus enabling more rapid development of future adapters.

- **Enhanced Parameter Mapping** – The proof of concept includes basic parameter mapping where all algorithms are applied to every data point in a file. However, we identified examples of files that would not map correctly to this basic structure. For example, if a file has both snow depth and temperature readings in one file, algorithms in the parameter mapping file applies to all data. This would not make sense as height measurements do not convert with the same multiples as temperatures. Therefore, Consortium members identified the need to handle parameter mapping algorithms per type of parameter. This would require the identification of a “parent” parameter within a group of parameters for a given data point. In the enhanced parameter mapping file, each algorithm could then be mapped to one of the “parent” parameters and would only apply to data within that parent structure.

- **Standard Hydrologic Dictionary** – As hydrologic data exchange increases and current barriers are broken down by these common data adapters, a next step for the Consortium could be to define a dictionary of common terms. This dictionary would define the HydroXC-recognized terminology and abbreviations for various hydrologic objects. For example, “river height” may be abbreviated “RH”. Organizations who adopt these standards further simplify their ability to share data with other organizations who have adopted them, as that removes another layer of custom mapping for each data exchange.

Community Building

Community building among the many participants in the hydrology field is of vital importance, especially at the practical level of data exchange and integration. Increasingly, organizations and agencies collaborate with each other across their boundaries and systems, exchanging data in numerous ways to produce and improve the necessary end products. All too often that exchange is cumbersome and difficult to set up due to the varying formats of data. HydroXC intends to further collaboration in the hydrology field by stimulating discussion among participants regarding data exchange around the common HydroXC format.

In order to bring the members of the consortium together regularly and keep everyone up-to-date on the latest HydroXC happenings, we carried out the following activities:

Website Upgrade

Until now, the HydroXC website existed as a subset of pages on the OHD NOAA website. This evolved simply from the fact that NOAA has been the sponsor of the HydroXC work, but the consortium is not a part of NOAA or OHD. In an effort to expand the capabilities of the site and enable it to live separately of NOAA and OHD, we ported the website to its own domain, HydroXC.org. The new site was given a unique visual design and improved navigation structure, which will allow for extended website capabilities in the future. The new website was launched in March 2007 and can be accessed at http://www.hydroxc.org. All of the content from the old site has been moved to the new site and all of the Phase 3 materials have been added. The website will continue to be minimally maintained through March 2008 and is being hosted at the CrystalTech hosting facility (www.crystaltech.com), paid through June 2008.

Partnership with Michael Piasecki, PhD, and Hydrology-related Conference

As described earlier, Michael Piasecki has been playing an important lead technical role for this phase of the Consortium work. Aside from attending all meetings and carrying conversations regarding the hydrological impact of HydroXC's mission, he presented the latest schema version 3.0 and hydrologic object representations at the *Geoinformatics 2007 - Data to Knowledge Conference* in San Diego, California, on May 17 and 18, 2007. The work was presented in a poster entitled, “A Common Schema for Hydrologic Data Transfer & Object Descriptions”. This poster can be found on the Continuing Work section of the www.hydroxc.org website.

According to Dr. Piasecki, the poster and HydroXC work garnered interest from many of the attendees. Interest ranged from generic inquiries about the consortium to more detailed conversations about the purpose of the schema. This conference was an important step in beginning to spread the word about the work being done with HydroXC and already garnered international inquiries and responses, i.e from the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) as well as requests to present the HydroXC work on symposia and other meetings. The HdyroXC work will be presented at the AGU meeting in San Francisco (December) as well as the Environmental Sensing Symposium in Boise (October).

Consortium Workshops

Much of the Phase 3 consortium work was completed either by Dr. Piasecki and his graduate students or the Apex team. In order to keep the rest of the Consortium up to date on our progress and to continue to grow interest from outside parties, Apex organized two different workshop meetings for all HydroXC members. These were timed to follow the main milestones of the work, as follows:

- **HydroXC XML Schema (March 22 and 23, 2007)**: During this workshop Dr. Piasecki shared his developments on the specific hydrologic object representations and the resulting schema. He introduced the channel cross-section, rating curve, and river reach object representations and discussed the need for the HydroXC schema to adopt a common geo-referencing format. He also displayed the new .xsd schema file and answered several questions from those on the call. All consortium members were encouraged to participate and discuss the progress.
- **Proof of Concept Data Adapter (June 21, 2007)**: During this workshop Glen Oliff and Stephanie Liu-Barnes, from Apex, demonstrated the proof of concept SHEF – HydroXC data adapter. After explaining the rationale behind the selection of the proof of concept format and describing the design, Glen demonstrated the capabilities of the adapter in real-time, including the following scenarios:
    - Conversion of basic unpacked .B SHEF message to HydroXC XML file
    - Conversion of basic HydroXC XML file back to a SHEF file
    - Location mapping to and from internal location identifiers to the corresponding latitude/longitude data
        - For example, “CMCN5” maps to latitude of 32.490 and longitude of -104.252.
    - Parameter mapping of name/value pairs
        - For example, map probability parameter “H” to “.75”
    - Parameter algorithm handling for value mapping (i.e. invocation of the conversion function instead of mapping the file lookup)
        - For example, convert parameter snow depth in millimeters to inches.

    We concluded the call with open questions and feedback from the attending Consortium members and also discussed options for finding new support for HydroXC going forward, outside of NOAA. Michael Piasecki works closely with CUAHSI and agreed to present the idea to the CUAHSI team.

Newsletters

Above and beyond the other contact points with the consortium members, the Apex team also e-mailed monthly newsletters summarizing the progress of the past month and reminding of upcoming events. Starting in September 2007 the newsletters will occur only every quarter. The final newsletter of this phase will be sent in March 2008.

Next Steps

Phase 3 is coming to a close with the first real and tangible artifacts of consortium work – the proof of concept data adapter and object representations. For the first time, members actually have something they can take away, use, and improve. For the next phase of the HydroXC we propose to continue to work on each of the overall themes listed above, as follows:

- **Research**
  - Continue to expand the object representations available in HydroXC to build a larger set that can garner momentum with the hydrologic community.
- **Development**
  - Identify a "real" project going on between two organizations and embed the HydroXC adapter work into its implementation. Members will only find the time and resources to try HydroXC and create the appropriate adapters when they have a specific work task identified.
- **Community Building**
  - Continue to hold open, periodic conference calls or meetings to present recent findings of the Consortium, discuss topics of interest, and demonstrate specific uses of XML in the hydrologic community.
  - Expand the website to include a discussion area, self-serve membership management, and more flexible options for others to contribute to its content.

Another important aspect for the HydroXC work is to find a more permanent "home" and support structure. NOAA began supporting this work as a project to determine the interest level across the hydrologic community and begin steps toward a standardized hydrologic data format. However, they never meant to fully influence and support the consortium work. The time has come for HydroXC to stand on its own more, possibly with the help of other efforts such as CUAHSI (http://www.cuahsi.org/his). Dr. Piasecki continues to explore this option and will present the idea to CUAHSI in the third quarter of 2007.

Conclusion

The HydroXC work to date has been productive and enlightening, resulting in new modeling concepts, software, and broad participation. However, the effort is still spearheaded and funded by OHD, and managed by Apex as a consulting firm, not a bona-fide participant who works in the hydrology community. Michael Piasecki has conducted early conversations with other organizations in the hydro modeling space, with the hope that an academic group or trade association may wish to further the practical, tangible results achieved so far. To achieve such an integration with other efforts, OHD may wish to support this work further, but focus more explicitly on transitioning the knowledge developed so far to other participants.