Extensible Markup Language as a Medium for Exchanging Hydrologic Basin Model Data

Andy Carter, E.I.T.[1], Trevor N. Barkhouse[2]

Abstract

In developing a hydrologic model for use in the National Flood Insurance Program, FEMA allows for the use and submission of over a dozen different software programs. These programs, created by various Federal agencies (USDA, USACE, and EPA), international groups (DHI) and recently commercial vendors (Haestad) have file structures for hydrologic basin models that radically differ from one another. Use of an existing electronic hydrologic model is currently dependant on having access to and knowledge of the software package the model was created in.

With easier access and distribution of hydrologic models to be accelerated by the creation and dissemination of the Enhanced Digital Flood Insurance Rate Maps, a standard file format is necessary to insure the shelf life and interoperability of a developed model. A standard modeling file structure must be created to ease translation between hydrologic program formats and allow use of legacy data.

Taking a cue from the technology sector, Extensible Markup Language (XML) seems to be perfectly suited for the task. A cousin of Hypertext Markup Language (HTML), XML provides many features necessary to be the Hydrologic model interchange format. XML is self-describing; text-based, and works seamlessly upon existing Internet protocols. Two of its companion technologies, Extensible Stylesheet Language Transformations (XSLT) and XML Path Language (XPath) allow for incredibly powerful data transformation and selection functionality. By defining the appropriate hydrologic basin data schema, a new standard can be established that various engineering tool providers can support. The common format would then allow for the exchange of data.

[1] Water Resources Engineer, Wier & Associates, Inc., 4300 Beltway Place, Suite 130, Arlington, Texas 76018, Phone: 817-467-7700, FAX: 817-467-7713; Email: andyc@wierassociates.com
[2] Software Developer, Usability Sciences Corporation, 909 Hidden Ridge, Suite 575, Irving, Texas 75038, Phone: 972-550-1599, FAX: 972-550-9148; Email: tbarkhouse@usabilitysciences.com

Introduction

For the hydrologist, technology is both a blessing and a curse. Just as it increases the productivity of our work, so does it foster our reliance upon it. The pace of the development of technology is so great that few can keep up with the latest advancements. This leads to one of the fundamental problems with computerizing information, which is the inevitable obsolescence of storage formats and mediums. In just four decades, we have seen the progression from punch cards, through magnetic media, and up to optical storage and flash memory. These too will become outdated. The bigger challenge, though, is the change in data formats. All computer data is just a long stream of numbers. Individual programs are designed to interpret the different bits as text, numbers, images, or other data. Attempting to access information that has been encoded in an unknown or incompatible format is nearly impossible. This problem is compounded by the differences between the popular computer operating systems.

Fortunately, as technology improves, consideration is given to the above-mentioned problems. Care is now given to changing data formats. Conversion programs are created and new programs are given the ability to open previous data formats. The most promising development is the drive for international standards. Compelled by several consortiums of influential technology companies, work is being done to overcome the differences in data formats to extend the useful life of computer data and to facilitate information exchange.

The field of hydraulic and hydrologic engineering is behind the cusp in this regard. Our hydrologic basin data is fragmented and in aging data formats. To understand the true nature of the problem, we must consider the impact of the National Flood Insurance Program (NFIP) on hydrologic modeling technology.

History of the NFIA

In 1968, the National Flood Insurance Act (NFIA) was passed by Congress, which authorized the issuance of federal flood insurance to communities that met floodplain-mapping requirements. To quantify risk and establish appropriate insurance rates, flood hazards were to be defined across the country. The Federal Emergency Management Agency (FEMA) was to oversee and maintain the floodplain-mapping used to calculate flood insurance rates.

At the time of the passing of the NFIA, engineering calculations for steady state open channel hydraulics and surface hydrology had been well established. However, with computational sciences in its infancy, computer programs to be used in determining flood hazard mapping were just beginning to emerge. The NFIA did not specify the methodology to be used in the determination of flood hazards. The following was the only directive that was given:

> "…Federal agencies engaged in the identification or delineation of flood-risk zones within the several States shall, in consultation with the Director, give the highest practicable priority in the allocation of available manpower and other available resources to the identification and mapping of flood

hazard areas and flood-risk zones, in order to assist the Director to meet the deadline established by this section."[3]

The Soil Conservation Service (SCS), United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), and the US Army Corps of Engineers each fell under this directorate. Prior to 1968, some of these federal agencies had already independently developed software for hydrologic analysis. Resulting from the lack of delegation in the congressional mandate, each involved federal agency either began to develop their own computer programs or continued to use those that had been developed at that time.

Modeling Philosophy and Methodology

Over the last 40 years, scores of surface hydrology modeling programs have been developed by a variety of groups for use in the NFIP. These groups include governmental agencies (both federal and local), private companies, and international organizations. Input files for these models have vastly different file structures.

So why were so many software packages, and therefore file formats, used for hydrologic modeling for the NFIP? The answer is twofold. First, when each group developed their hydrology modeling software, they were biased by their hydrologic modeling philosophy. Some federal agencies' programs incorporated only hydrologic algorithms that they had developed internally. Other packages were developed for solving only specific problems, such as runoff from small urban watersheds. Due to differences in modeling philosophy, cooperative development between the creators of the various hydrologic software packages has been minimal.

The second reason for the multitude of software packages is a function of the differences in modeling methodology. Computational technology available at the time that the software program was developed contributes heavily to the architecture of the file structure and computational limitations. Some packages restrict the number of sub-basins in the model, which was caused by a lack of computing resources. Over time, the increase in computational techniques has allowed for the incorporation of additional functionality and analysis methods into the more modern modeling programs.

The problems with the various data formats are not unique to hydrologic data and the field of civil engineering. Such hardship has been faced by every other industry that utilizes computers. By looking at the solutions to similar problems in the technology sector, we can customize the approach for hydrologic data.

Data Formats

Computer storage spans several types of media and countless formats. Initially, there were punch cards. The holes in the card stock would store data for a particular program on a specific type of computer. There really was not any concept of data exchange. Gradually, magnetic media became the de facto standard. Using magnetic tape, floppy disks, or hard disks, data could be archived and moved between different machines. Although hard disks are still in wide use, the main type of media for archiving data is now optical disks. All of these types of media store binary information (data represented by a series of bits).

[3] "Identification of flood-prone areas" Title 42 U.S.C. § 4101

To assist in consistency when exchanging text information, several standards were established for encoding the binary data in uniform ways. The most common were ANSI's American Standard Code for Information Interchange (ASCII) and IBM's Extended Binary Coded Decimal Interchange Code (EBCDIC). Through the use of these standards, two computer users could exchange text in the English language with some degree of certainty that the information would be correctly interpreted.

In 1991, an international consortium of technology companies[4] developed a new encoding standard known as Unicode. This encoding standard has been used commonly for the last decade. For encoding future hydrologic data, the use of the Unicode standard ensures that the text data will be accessible by more programs and for a longer time and allows for portability on an international level.

Federal Submittals in Proprietary Formats

Since April 2003, FEMA has allowed the submission of ten different types of single-event hydrologic models[5]. Most of these packages produce hydrologic input files that are digitally encoded in the ASCII format (such as HEC-1 and TR-55). There are also a few proprietary packages that produce files in binary, making the data less accessible.

Even though the five most commonly submitted hydrology modeling formats are from federal software packages, it is now possible to submit models in a proprietary format. In order for such a format to be accepted, the creator must grant permission to FEMA to release the code to impacted parties. Unfortunately, obtaining the proprietary software or source code from FEMA is a daunting task. Allowing the submission of hydrologic files in proprietary formats is detrimental to the public's interest. Review and use of this file is dependant upon access to the software package in which it was created. In the case of a proprietary format, often the only way to evaluate the input file is to purchase the software that created it. Then, even when the software is available, there is still expense and delay for developing expertise in the new program.

Continued Evolution of Modeling

Inevitably, hydrologic modeling philosophies and computational methods will evolve. Foreseeable advances include the creation of more finite analysis of basins. As computational methods improve, resolution of models will increase. Raster routing analysis will likely give birth to new empirical and theoretical hydrologic routing algorithms and software packages for creating surface hydrology models. For creators of hydrologic modeling software, each advance in methodology or philosophy creates a new file format, adding to the already long list that has been submitted to FEMA.

Universal Format

In a paper presented at the Second Federal Interagency Hydrologic Modeling Conference in the fall of 2002, Dudley McFadden, PE of Davis Ford Consulting Engineers proposed the

[4] "The Unicode Consortium" <http://www.unicode.org/consortium/consort.html>
[5] "Hydrologic Models Accepted by FEMA for NFIP Usage" <http://www.fema.gov/fhm/en_hydro.shtm>

development of a standardized hydrologic data exchange format[6]. Developing a standard file structure, he argued, would solve many of the problems created by the multitude of hydrologic model formats. The creation of a non-proprietary universal hydrologic modeling file structure would provide a neutral means to exchange, archive, and reuse data independent of the software that created it and allow for faster translation between formats. Developers of hydrologic modeling software would have a blue print for creating software that would conform to that file structure. Use of legacy models and models created in proprietary software would no longer be dependant upon the software package that created it. Furthermore, a common file structure would allow for the automated review of all submitted files.

To adapt to the continuous change of both modeling philosophy and methodology, a universal file format should be extensible. This would allow the standard file structure to be revised to add new techniques and capabilities as hydrologic science evolves without invalidating the initial universal file structure.

XML Solution

As McFadden[6] discovered, the obvious format for creating a standardized hydrologic modeling file structure was via Extensible Markup Language (XML). XML could be used as a medium for archiving and exchanging all modeling data prepared and submitted for use in the NFIP.

The evolution of the Internet has brought the levels of data collaboration to a new height. Part of the reason was the reliance of the World Wide Web on Hypertext Markup Language (HTML). This means of encoding information allowed for people using different types of computers with different operating systems to interpret information in a uniform way.

XML is a cousin of HTML. The difference lies in the grammar of the two. The World Wide Web Consortium (W3C) defines HTML. As of version 4.0, there are 91 different HTML tags[7]. These tags perform well when representing data for visual presentation in a web browser. However, they are not useful in representing data in any other way. Extensible Markup Language is similar to HTML in that it is a text-based format that uses tags to provide information about parts of the document. The list of tags is not set by a single organization, but is instead agreed upon by people wishing to exchange a specific type of information. For example, there is MathML[8] for mathematical formulae, CML (Chemical Markup Language)[9] for describing molecules, and OFX[10] (Open Financial Exchange) for banking data.

Another strength of XML is that it is "self-documenting". The format is comprehensible on its own, without additional instructions or documentation. For example, Figure 1 is a fragment of an XML document to describe a car. Even to the uninitiated, the meaning of the information is quite clear.

---

[6] McFadden, Dudley, PE "A Vision for a Standardized Hydrologic Data Exchange Format." Second Federal Interagency Hydrologic Modeling Conference, 2002
[7] "Index of Elements" <http://www.w3.org/TR/REC-html40/index/elements.html>
[8] "W3C Math Home" <http://www.w3.org/Math/>
[9] "Chemical Markup Language" <http://www.xml-cml.org/>
[10] "Open Financial Exchange" <http://www.ofx.net/>

```
<automobile type="Car">                <!-- Start of complex type -->
    <make>Hyundai</make>
    <model year="2000">Accent</model>
    <appearance>                       <!-- Start of complex sub-type -->
        <doors>4</doors>
        <color>Silver</color>
    </appearance>                      <!-- End of complex sub-type -->
    <options>                           <!-- Start of complex sub-type -->
        <radio>AM FM</radio>
        <powerWindows></powerWindows>  <!-- Indication of lack of feature -->
        <powerLocks/>                  <!-- Indication of lack of feature -->
    </options>                         <!-- End of complex sub-type -->
    ...
</automobile>                          <!-- End of complex type -->
```

Figure 1 – Sample XML describing an automobile

Over the last five years many engineers have developed and embraced XML as a medium for exchanging industry specific design data. For the land development business, a consortium of over 130 private and governmental organizations was formed to establish an open information exchange standard for surveying and land development data (LandXML[11]). For the geospatial industry, the OpenGIS[12] Consortium has developed the Geographical Markup Language for encoding geographical data. These are just two of the many of industry specific XML standards that have currently been defined.

HydrologicXML

Open any ACSII encoded hydrologic input file (SWMM, HEC-1, or HEC-HMS) in a standard text editor and you will find structured, detailed data describing the engineering parameters of the basin model. To the novice, interpreting a HEC-1 or SWMM model is difficult. Often a detailed reference manual is needed to translate the structure, methods, and input parameters. Without a corresponding work map, spatial reconstruction of this file is difficult.

Therefore, we propose the creation of a HydrologicXML format to universally describe all hydrologic basin data. Creating a HydrologicXML standard will provide a method for describing the complex structure of hydrologic data in a format that is free of licensing restrictions and is platform independent. Additionally, unlike the popular formats such as HEC-1 and SWMM, HydrologicXML would be both machine-readable and human-intelligible.

Many of the modern hydrologic programs produce modeling data that is fractionalized into multiple files. As with HEC-HMS, reconstructing a basin model requires many files (basin, DSS, DCS, and map). Implementing HydrologicXML would be flexible enough to contain all of the pertinent modeling information in a single file.

To convert an existing basin model from a current format to HydrologicXML a conversion program will have to be written for each specific software package. While it is best if

[11] "LandXML" <http://landxml.org/>
[12] "Open GIS Consortium, Inc." <http://www.opengis.org/>

the agency that created the original package write the translator, conversion software could be prepared by a third party. Creating a mechanism for converting to and from the universal format may seem like a costly approach, but it is preferable to writing converters between each data format. To achieve the same interoperability without a universal format, for *n* hydrologic modeling formats, $n(n-1)$ conversion programs would need to be authored instead of $(n * 2)$. See Figures 2 and 3 for a graphical representation of the different approaches.

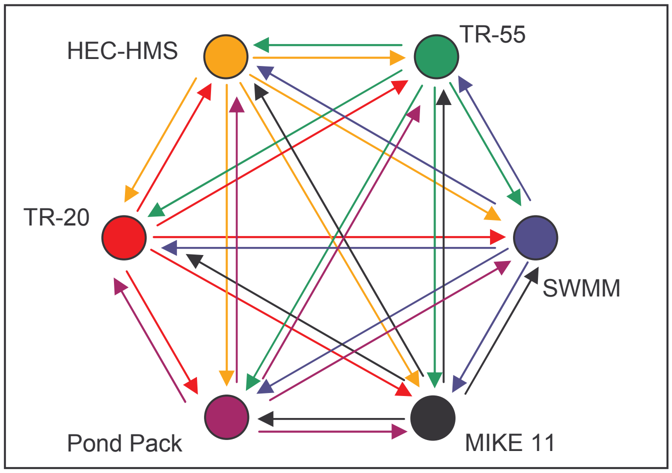

Figure 2 – The number of required conversion programs in the absence of a universal file format

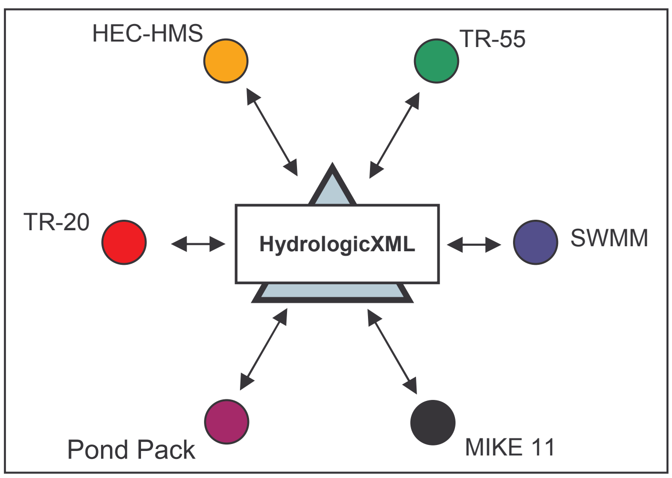

Figure 3 – The number of required conversion programs with a universal file format

Figure 4 contains a fragment of our sample HydrologicXML. This particular sample provides a definition of a sub-basin. Refer to our web site[13] for a comparison of the same data as it is represented in several other hydrologic modeling formats.

[13] "Additional Resources" <http://www.WierAssociates.com/HydrologicXML>

```xml
<basin name="C3">
    <position type="Single Point">
        <x>841.019</x>
        <y>1240.291</y>
    </position>
    <area>0.25</area>
    <downstreamElement name="Junction-1" type="Junction"/>
    <lossRate method="SCS">
        <percentImperviousArea>0.0</percentImperviousArea>
        <curveNumber>78.5</curveNumber>
        <initialAbstraction>1.25</initialAbstraction>
    </lossRate>
    <transform method="Modified Clark">
        <timeOfConcentration>0.52</timeOfConcentration>
        <storageCoefficient>0.25</storageCoefficient>
    </transform>
    <baseFlow/>
</basin>
```

Figure 4 - Excerpt of sample HydrologicXML

XML Schema

An XML schema is a definition of the proper syntax and grammar for an XML document that contains a specific type of data. For example, the schema for LandXML defines the elements and their proper ranges for survey and land development data, such as alignments, survey points, and surfaces. The schema serves three main purposes. First, it is a formal agreement about the structure of the data. Schemas are published so that different parties are able to exchange information in a uniform manner. By conforming to the schema, two software packages can provide information with each other without the creators ever having to interact. Second, the schema serves as documentation for the XML dialect. By looking through the schema, a person can get an idea about the possible types and values of data for the given dialect. There are also "annotation" elements in the schema that allow the author to insert notes for other viewers, be they people or applications. Finally, the schema can be used by applications to verify that the given XML is valid; that is that it contains all required information and the values are in the appropriate ranges. A program can use the schema to ensure that the data being provided by an unknown (and thus un-trusted) source is properly formed. Technically, XML does not require a schema. However, they are invaluable tools to ensure compliance with a standard. Such standards are necessary as information is more broadly disseminated.

The scope of the HydrologicXML schema definition as defined herein was limited to just the basin modeling. Ultimately a complete hydrologic model schema would include additional items such as meteorological information, gage data, and calibration configurations. However, this definition is limited only to defining the physical properties of the spatial relation of a dendritic hydrologic model. Specifically, the only three elements included in this initial definition are basins, junctions, and reaches. Future work will include the incorporation of reservoirs, diversions, sinks, and other pertinent elements. See Figure 5 for an excerpt of the sample XML schema that we created for our proposed HydrologicXML, which describes a junction.

```
<xs:element name="junction" minOccurs="0" maxOccurs="unbounded">
    <xs:complexType>
        <xs:sequence>
            <xs:element name="description" type="xs:string" minOccurs="0" />
            <xs:element name="position" type="position />
            <xs:element name="downstreamElement" type="downstreamElement" />
        </xs:sequence>
        <xs:attribute name="name" type="xs:string" />
    </xs:complexType>
</xs:element>
```

Figure 5 - Excerpt of schema for HydrologicXML

To implement HydrologicXML across the industry, a schema must be defined to establish rules for uniformly describing hydrologic data. We have provided our HydrologicXML schema definition as a foundation upon which the hydrologic industry can build. Ideally, a working schema specification would be developed through the cooperation and participation of federal agencies, software developers and private firms. FEMA may be best suited for sponsoring an industry wide collaboration, and by doing so could provide the necessary influence for integrating the new HydrologicXML standard.

Extensible Stylesheet Language Transformations (XSLT)

The XML suite of technologies contains a means of converting XML data into other formats. The terminology for such a conversion is "transformation". Common transformations are from XML to HTML for viewing news articles on the Internet through a web browser. The transformation of data is one of the most import aspects of any data format. Too often, it requires custom programming to get data from one format to another. XSLT provides an existing technology for doing so. By having several different transformations, the same XML data can be converted into several different forms. This is demonstrated by transforming our hydrologic XML samples into HEC1, HEC-HMS, and WinTR-55 formats, as demonstrated on our web site. Another use is creating user-friendly reports with HTML. These can do everything from summarizing data to illustrating invalid data. Figure 6 contains a graphical representation of the various transformations.

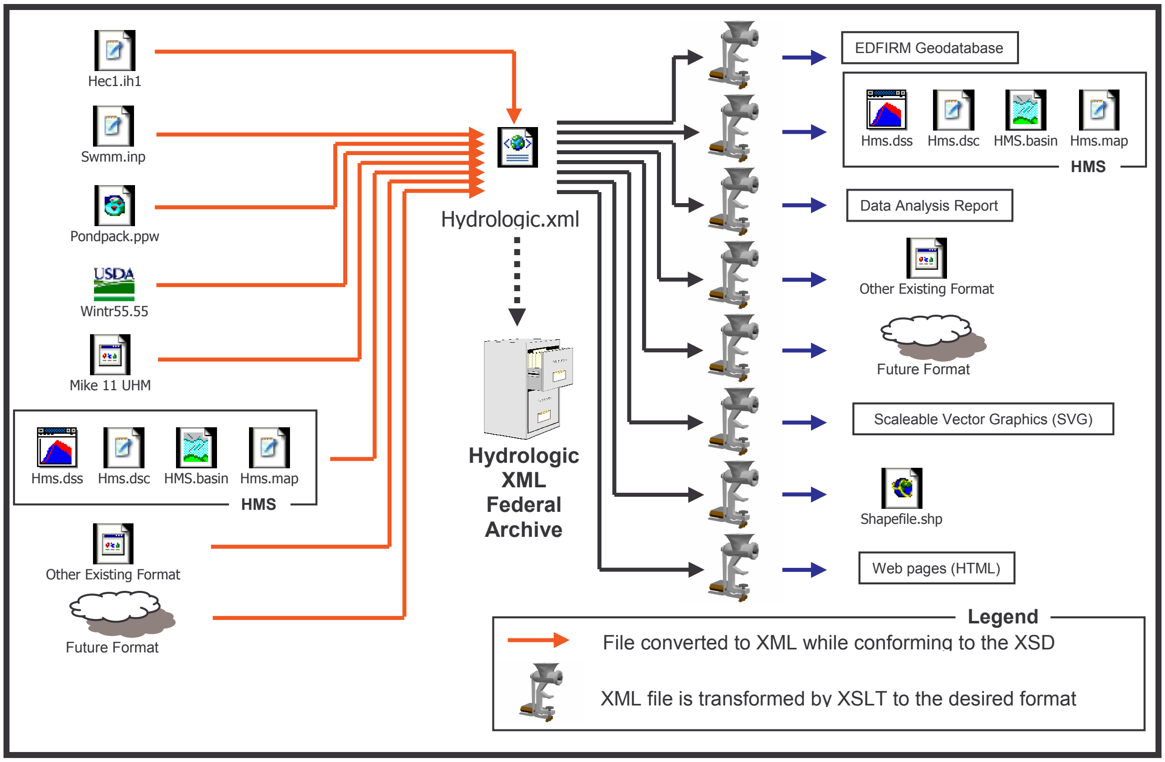

Figure 6 - Transformations between hydrologic modeling formats with HydrologicXML

If a HydrologicXML file is attributed with geospatial data such as element position and sub-basin boundaries, an XSLT graphic translation could be provided to convert these properties into a variety of vector graphic formats. This translation could be to proprietary formats such as ESRI shapefiles, DWG's and DGN's. Furthermore, internet based vector formats (SVG, VML) could be created via an XSLT to provide the hydrologist with a plan view of the basin model in a web browser.

National Digital Mapping Data

In 1997 FEMA began the process to update and replace all of its paper flood insurance rate maps with digital mapping. The paper flood insurance rate map (FIRM) will be replaced with a digital version called a DFIRM. The DFIRM contains geospatial vector data and raster aerial imagery showing the effective mapping. Additionally, each DFIRM panel contains a geospatial relational database containing all of the pertinent data in reference to the floodplain mapping.

In February of 2002, FEMA expanded the standard DFIRM database definition to include engineering data and modeling used in floodplain definition. The enhanced DFIRM database (EDFIRM) includes a compressed copy of the hydrologic model. This methodology could be improved by attaching a hydrologic model in an industry accepted XML format to provide a greater insurance of future interoperability. With the option of providing geospatial attributes in the defined schema, population of the hydrologic portion of the federal relational databases for an EDFIRM could be preformed by the use of an XSLT.

The Use of XPath and XQuery to Query Hydrologic Information

The specifications for XML Path Language (XPath) and XML Query (XQuery) are related to the XML specification. These two technologies allow for the selection and navigation of XML data that fit certain criteria. For example, it would be possible to select all of the sub-basins that shared a common junction in a hydrologic basin. On a larger scale, it would be possible to select all of the models that were prepared for a specific region (city, county, even state). This functionality does not need to be programmed independently, as it is part of the standard that governs XML and is uniformly implemented by any program that uses XML.

Concurrent to the creation of the EDIFRM database, the creation of a hydrologic model federal archive is recommended. This archive would be a national repository for the storage of all hydrologic models in a HydrologicXML format. From this archive both text-based and spatial queries could be made using internet protocols to locate all models that include a specific area.

Conclusion

The creation of a non-proprietary, industry-wide standard for describing hydrologic data is imminent. Employing XML as the medium for creating a standard file format, allows the hydrologic engineering community to leverage many of the existing XML technologies as a means for increasing productivity and accuracy. By storing our data in an XML format, hydrologic models would be human intelligible while remaining machine-readable. An extensible universal hydrologic format would provide an effective method for archiving existing and legacy models while providing flexibility to incorporate future industry modeling advancements.

For the hydrologic modeling industry to embrace HydrologicXML as a means for describing and exchanging basin data, a schema definition must be refined to account for the many methodologies and philosophies of current modeling programs. While we have defined an initial basin schema, the working specification will evolve as a result of industry collaboration. Upon the foundation of a formalized HydrologicXML schema specification, hydrologists will have a medium for exchanging hydrologic modeling data.

Additional Resources

Refer to www.WierAssociates.com/HydrologicXML for additional information regarding HydrologicXML. This site contains the initial schema definition as well as sample HydrologicXML files and transformations (XSLT).