ABSTRACT
The Indian Ocean Tsunami of 26 December 2004 made it clear that information about sea-level stations that could be used to support detection and warning (such as location, collection and transmission capabilities, operator identification) are insufficiently known or not readily accessible. In addressing this issue, agencies within the Pacific region are collaborating to develop a web service to expose station metadata enabling various types of real-time data mining client applications that support decision making and strategic planning at Tsunami Warning Centers (TWC).

Categories and Subject Descriptors
H.2.8 [Database Applications] Data mining, Image databases, Scientific databases, Spatial databases and GIS, Statistical databases.
I.7.2 [XML] Desktop publishing, Format and notation, Hypertext/hypermedia, Index generation, Languages and systems, Markup languages, Multi/mixed media, Photocomposition, typesetting, Scripting languages, Standards.

General Terms
Management, Measurement, Documentation, Performance, Design, Standardization, Languages.

Keywords
Data mining, Knowledge discovery, Disaster management, Civil defense, Web service, Geographic Information System, Warning center, Coastal hazard, Tsunami, Sea level station, Tide gauge, Sensor, Wave water level, Spatial database, Publish subscribe, Service Oriented Architecture, Business Intelligence, Remote Procedure Calls, Framework, Pacific, Hawaii, XML, GIS, SOA, SOAP, HTTP, BI, PDC, PTWC, HI-RISC, NOAA, IDEA, WWL, Google Earth, Tide Tool.

1. INTRODUCTION
The Tsunami of 26 December 2004 exposed a grave need for sea-level station data to be at the fingertips of warning center personnel and disaster managers. Specifically, in the critical moments after the earthquake was reported, it was unclear to officials if there were any sea-level stations located nearby operating in real-time and, if so, whom to contact to obtain the water level records. In post-disaster hindsight, it became obvious that information about sea-level stations that could be used to support detection and warning (such as location, collection and transmission capabilities, operator identification) are insufficiently known or simply inaccessible.

The first order of business for agencies united under the Pacific Region Data Integrated Data Enterprise (PRIDE) program by the National Oceanic and Atmospheric Administration (NOAA) was to develop a plain language semantic description of a sea-level station as applicable to tsunami hazards. This semantic description was later formalized into an XML schema definition.

Thus began, in the aftermath of the Indian Ocean event, a multiyear effort to develop a distributed metadata system describing sea-level stations starting with pilot activities in a regional framework and focusing on tsunami detection and warning systems being developed by various agencies. This activity has matured at the time of writing (May 2007) to whereby metadata for stations monitored by the Pacific Tsunami Warning Center’s “Tide Tool” are now available for consumption by XML enabled client applications. Specifically, the service is available in two formats: KML, which can be opened with Google Earth, and XML. Through the service provided by the Pacific Disaster Center (PDC), users can visually locate sea level stations in an area of interest (such as in immediate proximity of an under sea earthquake) and simply click on station icon to access the Wave and Water Level (WWL) records at the station.

During the past couple of years, the data content and sharing issues embodied in this schema have been discussed at various forums. The result is that the various stakeholders have emerged who have different provider and user perspectives (semantic content) and also exchange formats (not limited to just XML). The challenge then, is not only to capture all data requirements, but also to have formal representation that is easily transformed into any specified format. The latest revision of the sea-level gauge schema (Version 0.3), begins to address this challenge. It encompasses a broader range of provider and user perspectives, such as station operators, warning system managers, disaster
managers, other marine hazard warning systems (such as storm surges and sea level change monitoring and research).

Users will be able to create custom client applications to quickly drill-down to detailed information or conduct custom transformations to monitor a real-time event using tools like the PTWC TideTool, and access the data streams as well.

Some of these transformations can be simple XML-to-XML transformation such as the KML standard supported by a rich client like Google Earth, as shown in Figure 1. Others can support loosely coupled SOAP web service clients applications that consume the same service for different purposes.

The gist of Figure 2 is that data sources can be of various types but the client side view is intended to be integrated and seamless thanks to the web service. The client could GIS-based, as shown. Or, it could be a SOAP-based client that extends the published interfaces to mine the web service for data according to the needs of various domains (such as station managers, warning system managers and emergency managers), as shown under user perspectives in Figure 2.

2. REPRESENTATIVE USE CASES

As shown in Figure 3, there are four chief methods of consuming the XML station metadata released by the PTWC/PDC:

- **Raw**: Various XML aware applications such as web browsers and sophisticated XML editors can be configured to visualize some or all of the station metadata being exposed.
- **Transformation**: By applying a stylesheet or equivalent programmatic transformations, the raw XML may be transformed into markup languages like KML/HTML to enable specific clients like Google Earth.
- **Digest**: Fundamentally, the digest pattern involves absorption or conversion into completely different object models or schemas. Typically, these object models are used to populate customized database tables (such as GIS layers).
- **Remote Procedure Calls (RPC)**: Various applications may build on the schema for the stations XML file and feature remote procedures to mine the data referred to by the metadata contained in the file.

A couple of representative use cases (client applications) are described below:

- Pacific Disaster Center (PDC) Hawaii Resource Information System for Coastal Hazards (HI-RISC)
- Pacific Tsunami Warning Center (PTWC) Tide Tool
### 2.1 HI-RISC

Hawaii Resource Information System for Coastal Hazards (HI-RISC) is a Geographic Information System (GIS) platform for integrating, applying and disseminating risk management-related data and products in a web-based mapping environment. With an emphasis on the development of integrated data products, incorporating model-based decision support tool outputs and near real-time observations with baseline GIS data, HI-RISC will support the creation of multi-hazard vulnerability maps and their use by decision makers in Hawaii for disaster risk reduction.

One of the numerous “layers” in the HI-RISC application is the Wave Water Level (WWL) layer featuring sea-level station metadata. In many ways, this client functionality is similar to Google Earth (visually). In fact, since this is a PDC hosted application they share the same object models and servlet designs (such as the balloon pop-up showing station metadata). However, the key advantage of HI-RISC over Google Earth is that the user can now overlay sea-level station data with numerous other custom layers pertinent to disaster management and conduct sophisticated GIS based queries. This is impossible in Google Earth.

Further, in the simple Google Earth (free) client, the user has no control over the base imagery. This is generally acceptable for the layman, but disaster managers often need time-sensitive imagery for disaster recovery (e.g. before and after the tsunami) or spectra other than the visible (such as a query result generated by a database-level foreign-key relationship with shelter occupancy) which only a custom GIS such as HI-RISC can provide.

### 2.2 Tide Tool

Stuart Weinstein developed Tide Tool as a platform-agnostic tool (Tcl/Tk environment) to process sea level data at the Pacific Tsunami Warning Center (PTWC). Sea-level data is transmitted across the GTS (WMO) circuit and flows to the PTWC over a dedicated circuit. The data is downloaded into a log file that is changed daily. PTWC’s Watertool program reads the log file every 15 seconds to retrieve new data. Watertool has three main parts: a decoder, a GUI and a monitor. Graphics are created by the Tcl/Tk programming language. Tcl/Tk is Open-Source and is platform/OS independent. While Tcl/Tk may be platform independent, Watertool is a complicated piece of software that is not very transportable. Watertool is an embedded Tcl/Tk application. In an effort to create something transportable and easy to maintain, Dr Weinstein developed Tide Tool as a pure Tcl/Tk application.

In the Pacific Ocean region, sea level gauges are maintained by various agencies such as PTWC, ATWC, NOS, SHOA, NTF, JMA, NDBC, CISESE and UHSLC. The data are received by satellite and forwarded to PTWC via the NWSTG. The DAPS system is used as a backup. The systems are being constantly upgraded. For instance, while many stations still transmit hourly, West Coast, Hawaii, Puerto Rico, and Alaska NOS gauges have been upgraded to a 6-minute transmission schedule. In any case, the decoding of raw sea-level data is fraught with breakdowns because several ever-changing data formats are being used.

#### Examples:

**UHSLC Format – Readable ASCII**

```
SWI040 RJTD 250015
:ENB 1 #1 M 3908 3908 3910 3909 3911 3909 3912 3910 3913 3913 3917 3917 3917 3917 3915 3918 3914 3917 3912 3913 3913 3912 3912 3911 3908 3908 3905 3909 :ENC 1 #2 3409 3410 3411 3411 3414 3413 3419 3419 3420 3419 3415 3414 3418 3411 3408 3410 3409 3409 3409 3414 3413 3409 3414 3414 3410 3412 3409 3410 3413 :BATTLOAD 0 12.83 :NAME=
```

Values are sea level in one hundreds of a foot. Newer gauges are metric.

**NTC Format - Unreadable Format**

```
91642 46/// /1205 10296 40080 22200 00287 555 77744 A0102 5163 60029 6315B 03024 83030 00A07 02548 02901 29631 6B090 24520 2400A 13025 90036 00297 317B1 52040 60310 0A190 26230 38002 96317 B2102 37103 100A2 50266 50330 02973 18B27 02331 02800 34152 02917 3419 3420 3419 3415 3414 3418 3411 3408 3410 3409 3409 3409 3414 3413 3409 3414 3414 3410 3412 3409 3410 3413 :BATTLEAD 0 12.83 :NAME=
```

**NOS Format - No Apparent Pattern**

```
206011307M94168411DZpQ0@@rf10uW@1Am0@-eBSBYBAG@BrBvBzAM@BqBZCUBA@BoBZCMbA@BmBZDMAI@BIBIZDrAG@BIBZD@AD@BkBZETAL@BkJBewAGABhB@0v@ol3@[DvAI4B]5Ad6a=-OE0vWv>ZAnAlBYBkCrCqCsDWDnER"@us@so0uVv>YCqC"\DnDxEGEa"@wh@wB _OLAoP 50+1NN 116W
```

Figure 4. Station metadata in HI-RISC Client
This represents an unpredictable requirement for Tide Tool to first figure out how the data is encoded and how frequently it is transmitted. In fact, a considerable amount of code in Tide Tool is dedicated to keep up with the changing station metadata from various agencies. With a sea-level station web service that is dedicated to keeping station metadata current, the Tide Tool code base can be stabilized and focused in its area of core competence which is to conduct data mining and to display water level data.

Specific actions we plan to take in the near term include:

- Recruiting station providers to actively provide, maintain and disseminate data through this shared web standard.
- Refining the XML-based tide station metadata schema by following the process that lead to development of the pilot sea-level station metadata web service. Consult with station operators and user community regarding additional elements in the XML schema. Ideally, expand it to include metadata applicable to other marine hazard warning systems (such as storm surges), sea level change monitoring and research.

Over the medium to long-term we will strive to:

- Make web services user-driven. Specifically, create data mining remote procedure calls enabling clients to maintain and harvest data for various purposes.
- Promote the regular use of web services in disaster training, simulation and exercises. Only through regular usage will we be able to work out the defects to a point where the web service enabled clients can be useful in a real event to save lives.
- Explore relevant and emerging standards, such as, the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) Framework, Tsunami Warning Markup Language (TWML) etc. that will serve all prospective stakeholders in the community most usefully (i.e. in an extensible, scalable) manner.
- Maintain the thrust in breaking new ground and establishing a standard format (specifically, a workable XML schema definition) for message products (e.g. tsunami bulletins). This would introduce the ability to adaptively parse legacy bulletins as-is (instead of imposing any changes to established processes).
- Fully integrate a GIS client like Google Earth or HIRISC with an HTTP enabled desktop client application like PTWC Tide Tool, display tsunami arrival times.

The development of a sea-level station metadata web service and associated activities described here is part of a broader vision of coastal inundation and erosion data and product development that:

- Effectively integrates observations, research, integrated modeling, forecasting, assessment, information management and education; and
- Involves a true partnership among the providers of information on coastal conditions and the users of that information in government, resource management, community planning, business and science (Marra, et al., 2007).

The desired outcome is that planners, managers, and other decision-makers will have timely access to information that is both accurate and appropriate, and, as such, will afford them the opportunity to plan and respond accordingly. As a result, the resilience and adaptive capacity of coastal communities affected by inundation in all its forms will be enhanced.
4. ACKNOWLEDGMENTS

NOAA IDEA Center, Eileen Shea, Director, leads various programs that have contributed to this paper. We received key inputs from UNESCO IOC/IODE staff, especially, Peter Pissierssens and Vladimir Vladymyrov. Mark Merrifield at the University of Hawaii Sea Level Center and Laura Kong at International Tsunami Information Center assembled the initial PRIDE web services project team. For cost-sharing labor and committing their agencies, we appreciate our respective bosses, both at the PDC, Ray Shirkhodai and Chris Chiesa as well as at the PTWC, Charles McCreary. We are grateful to Auroop Ganguly, ORNL, for creating this link between the broader KDD world and ours (disaster management). Finally, much thanks to our esteemed IT systems colleagues at the PDC and the PTWC who continue to make operational many innovations in disaster management.

5. REFERENCES


