Metadata Working Group User Requirements Committee
Use Case Interview– 7/30/02
 ====

Steve Heard, Director of Technology
Public Interactive
(617) 423-4499, sheard@publicinteractive.com

Interview Conducted by Tim Olson

Advice to Metadata Committee
   •   Don't make dictionary to complicated to use
   •   PI model is proof that a few, simple metadata fields works (80 stations
       now using PI system without complaint)
   •   Asking reporters/producers to input to more than a few fields won't work

Interview Highlights
   •   Uses XML to distribute stories, using the News ML format/standard
----

What Public Interactive Does
Public Interactive is an application service provider with approximately 80 public
broadcasting clients, primarily radio stations. They provide stations web tools
such as events calendars, story publishers, audio streaming and e-commerce
engines for stations to publish their content under their station Web site look &
feel. Public Interactive's primary constituency is stations, not end users.

Public Interactive Assets
   •   News
   •   Stories
   •   Audio
   •   Calendar Events

News-
PI imports national news feeds from Reuters, Screaming Media and others and
re-syndicates these to stations via XML and Javascript includes. For the XML
standard, PI uses NEWS ML. NEWS ML is a metadata set originated by
Reuters that is now a de-facto "standard" for the news/journalism field. NEWS
ML utilizes fields such as headline, dateline, byline, story, associated video and
associated audio. PI adds fields for teaser text and section (see below) before
sending it on to stations.   Of the 80 stations, only four can utilize the XML
version of the feeds due to lack of staff XML developers.

Stories -
PI publishing engine allows station users to upload stories/reports and a small set of associated metadata including headline, teaser, associated audio, section and keywords.

Keywords - PI keywords field has no controlled vocabulary, station users upload terms at their discretion. There is no spell check. Total keywords in the system is now over 10,000.   End users are provided list of related stories by PI engine running an automated same keywords search.

Keyword Search – users search the keywords field, not the entire story text field. Entire story text field search would be to slow.

Section – PI provides eight section categories originally, but station users are allowed to add others as they see fit. Many stations name new categories individual radio show names. PI original sections are - top stories, local news, US news, world, business, opinion, science, arts, sports, learning and Reuters. Reuters is a special category due to the volume of Reuters stories. Arts has sub sections of classical, Jazz, pop, books, TV, movies, people, programs, radio, theater, headlines and columns.

Audio – binary audio (the words of inside the audio itself) are not transcribed or indexed. Stations assign metadata to finished/complete audio pieces as stories. Fields include date published, tease (1/2 paragraph of text), dateline, byline, and keywords. Users searching audio stories are actually searching the keywords field in stories.

Audio is all completed segments/reports, primarily 4-5 minute pieces. However stations are increasingly posting shows in their entirety (30 or 60 minutes).

Calendar Events – Not tied/linked to anything else (stories, audio..). Events have start date, end date, venue, description, and classification. Classification is done at station or user discretion. Users are allowed to submit events directly and make up their own classification.

Future – In Audio Binary Search
Though few/no stations have been asking for it, PI is eventually moving to classify content of the audio itself, allowing users to search every word spoken during a particular speech or report. The main obstacle for implementation is cost-effective automated in-audio speech to text software.