

Environmental Health News -- Building an on-line news portal using Semantic Web technologies

Cormac Twomey
Siderean Software,
Los Angeles, California, U.S.A.
ctwomey@siderean.com

Todd Koym
Edgerton Foundation,
Washington, DC, U.S.A.
todd.koym@edgertonfoundation.org

John Peterson Myers
Environmental Health Sciences,
Charlottesville, Virginia, U.S.A.
jpmyers@ehsic.org

Abstract

Environmental Health News (EHN) is an on-line resource that covers issues concerning human health and the environment. Published daily by Environmental Health Sciences, its mission is to help increase public understanding of emerging scientific links between environmental exposures and human health. The EHN service provides a web site and delivers email newsletters to a growing audience of readers and subscribers, providing them with frequently updated summaries of breaking news, scientific research and reports by environmental organizations. All content is tagged with terms from the EHN subject taxonomy, which covers a broad spectrum of environmental health issues.

Keywords

Content management; Educational Web Applications; Faceted Navigation; Environmental Health; Taxonomy

1. Introduction

In July 2003, the Edgerton Foundation and Environmental Health Sciences contacted Siderean Software to investigate enhancing the EHN service by making use of Siderean's faceted navigation engine, Seamark. At the time, EHN was manually maintained, with about 30 to 40 article summaries added per day. This labor-intensive editorial process was an impediment to extending EHN's scope and audience. To achieve its ambition to be a primary resource for activists, scientists, media outlets and the public at large with respect to matters of environmental health, EHN needed an innovative content management system and information architecture that would allow it to grow. To address this, the Edgerton Foundation and Siderean Software jointly developed a new content management system built from the ground up using semantic web principles. The new EHN service encodes all content on the site in RDF, using standardized terms. Specifically, the new site was designed to:

- Aggregate content about health and the environment from distributed, heterogeneous sources.
- Support the creation and aggregation of value-added metadata about the content.
- Provide effective and flexible navigation and search over a large and growing repository of content metadata.
- Support the creation of customized syndicated feeds of the value-added metadata for delivery across diverse media platforms.

The new EHN service is a proprietary application designed by Siderean and EHN runs on a single RedHat Linux PC, hosted by Siderean. It is comprised of two main components:

- Apache Tomcat hosting Java Server Pages which provide the front-end user interface and editorial workflow.
- Siderean Seamark Server acting as the metadata store and navigation engine.

Both components run on the same machine, and communicate with each other via a SOAP interface. The current version of the EHN service can be visited at <http://www.environmentalhealthnews.org/>.

2. Navigational Features

2.1 Faceted Navigation

Environmental Health News: Compiled by Environmental Health Sciences - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.environmentalhealthnews.org/archives.jsp

Environmental Health News

Compiled by Environmental Health Sciences

Front Page Archives Subscribe Syndication About EHN Help

Search Go

Current query: 1 to 15 of 25209 items | [next](#)

Refine:

by Text Search

All fields Go

by Type

- [Editorials](#) 1099
- [News Stories](#) 22972
- [Opinion Pieces](#) 737
- [Organization Reports](#) 205
- [Scientific Studies](#) 197

by Current issues

- [Air](#) 6126
- [Cancer](#) 4287
- [Climate change](#) 2069
- [Energy](#) 9923
- [Environmental politics](#) 5238
- [Hazardous products](#) 4559
- [Pesticides](#) 2827
- [Sewage systems](#) 2830

While all links worked when entries were posted to the database, different publishers have different policies about retaining articles and providing access to archived material. Thus some of the links, particularly older ones, may no longer be functional. For links no longer working, you may be able to gain paid access to text via the publisher's site.

The more we grow, the less able we are to feed ourselves. The world is consistently failing to grow enough crops to feed itself, alarming official statistics show. Stored stockpiles have fallen sharply and are now at the lowest level on record. [London Independent](#), England. 29 August 2004. [\[related stories\]](#)

Demographic 'Bomb' May Only Go 'Pop!' The threat of a population bomb is now more regional than global, explosive only in places like India and Pakistan. [New York Times](#). 29 August 2004. [\[related story\]](#)

Canada Reinforces Its Disputed Claims in the Arctic. A \$4 million military exercise is the most prominent sign to date of Canada's intensifying effort to reinforce disputed claims over tens of thousands of miles of Arctic channels and tundra, whose accessibility is being altered by climate change. [New York Times](#). 29 August 2004. [\[related stories\]](#)

Pollution 'slows down' workers. For the first time, researchers have measured how sick building syndrome - the mystery flu-like illness often blamed on germs spread by office air conditioning - could be affecting workers' performance. [London Observer](#), England. 29 August 2004. [\[related stories\]](#)

A case study for 'sick buildings'. A state office building in Hartford could soon provide insight into the national mystery known as sick building syndrome. [Hartford Courant](#).

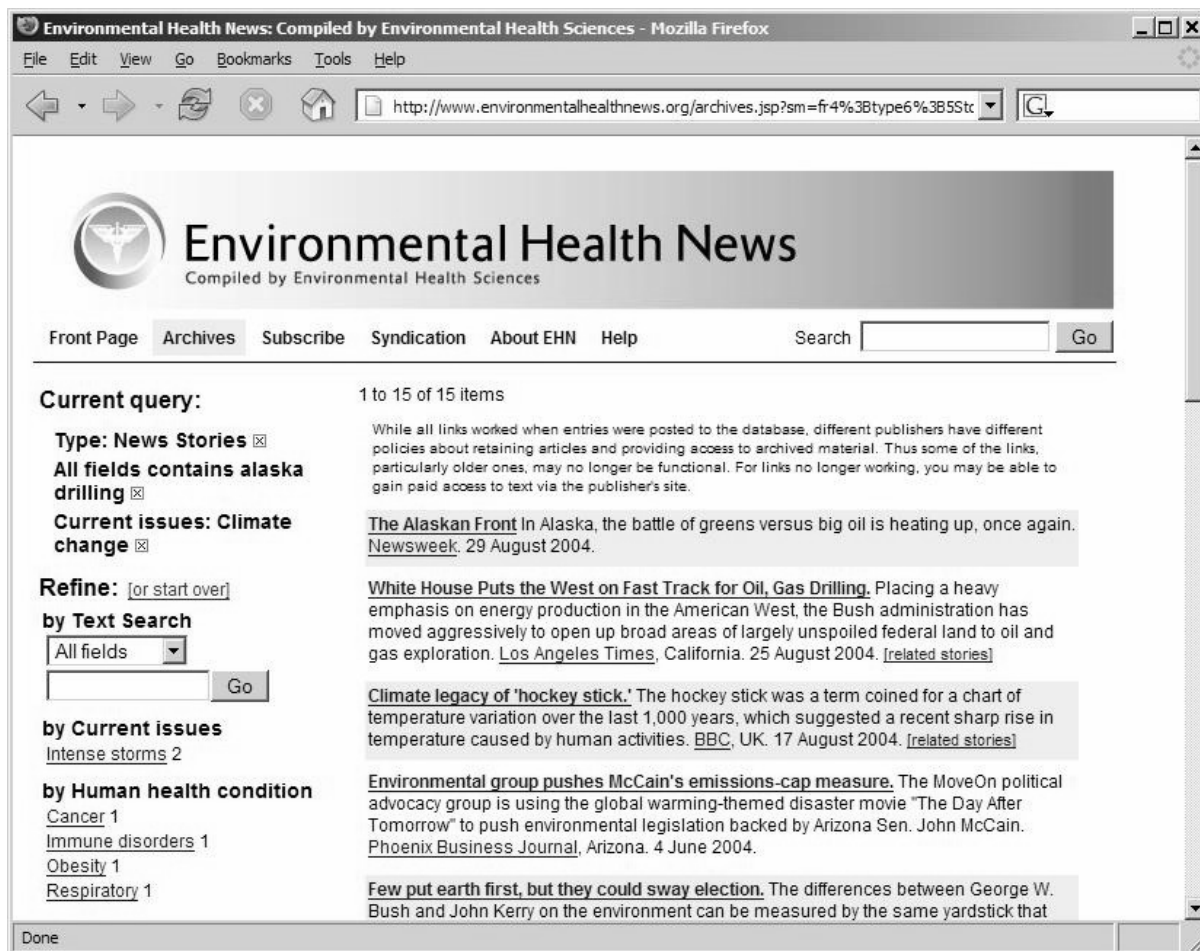
Done

Archives Page

All articles on the site have been classified according to the EHN taxonomy. This categorization has been leveraged throughout the site, in particular on the site archives page, which allows the user to browse the results of their search using faceted navigation [1]. The top-level terms are shown as separate navigation facets on the archives page:

- Human Health condition
- Contamination agent
- Exposure pathway
- Ecological effects
- Infrastructure
- Solutions
- Emerging Science

The article type and publisher information are also exposed as navigation facets.



Example Query

2.2 Customized Syndication

One of the primary goals of the EHN project is not only to provide a centralized repository of environmental health content, but also to ensure that it can be leveraged and reused by third parties. To this end, all content on the EHN site is available through RSS 1.0 feeds. Indeed, each individual search results page can be obtained as RSS 1.0, providing site visitors with customized feeds tailored to their specific interests. These feeds can then be used by individuals to monitor any area of particular interest. For example, users of the site can create RSS feeds to monitor:

- new articles that relate to a topic or set of topics in the EHN taxonomy;
- new articles added to the site that are related to a particular article of interest;
- new articles that correspond to a text search, or
- any combination of the above

2.3 JavaScript Syndication

EHN JavaScript Feed Wizard - Mozilla Firefox

This wizard will walk you through a few simple steps to help you match the formatting of the feed to that of your website.

In this first step choose how many news items you want to display, then specify the how you want the headlines to be formatted. Click next to go to the next step.

After the last step you will be given a snippet of Javascript that you can insert into your web page so you can display the feed on your own website with the formatting options you've chosen.

Click Finish anytime the feed looks the way you want it to.

Headlines

of items:

family:

size:

weight:

color:

CSS class:

[Cancel](#) [Next](#) [Finish](#)

Environmental Health News

Search

31 August **Battle over toxic metal.** Over the objections of several federal scientists, the Bush administration is preparing to relax national standards for selenium - a toxic metal that caused mass deformities of water fowl in California's Central Valley during the 1980s. [Sacramento Bee](#), California.

31 August **The fall of a pesticide policeman.** The troubled tenure of a state inspector provides a window into the state's efforts to safeguard human health and the environment as farm chemicals are sprayed in the spring and summer across Washington's farmlands. [Seattle Times](#), Washington.

31 August **The Man Who Knew Too Much.** A BU professor fights to publicize his report claiming there's a link between cancer deaths and IBM plants. But Big Blue is doing all it can to stop him. [Boston Magazine](#), Massachusetts.

Javascript feed

In addition to RSS syndication, EHN also offers JavaScript-based syndication, allowing websites to drop a feed from EHN directly onto their website simply by adding an external JavaScript link to their web page. The site offers a JavaScript feed wizard which allows the user to customize the appearance of the content to match the look & feel of their own site.

3. EDITORIAL FEATURES

3.1 Manual metadata markup

EHN editors submit new articles to the EHN site by using a simple HTTP bookmarklet when viewing the page to be submitted. This raises a window with a two pane view -- metadata on the left, article on the right:

TheStar.com - Getting the lead out' - Post to EHN - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

New Article Submission:
<http://www.t...l=969048863851>

Update URL:

Title:

Lede:

Description (2-3 sentences):

Reporter:

☒ Publisher: **Toronto Star**
 Publisher URL: <http://www.thestar.com>
[Show article](#)

Article Type:

TORONTO STAR

www.thestar.com
 Mon. Aug. 30, 2004. | Updated at 05:28 PM

Home

GTA

Business

Waymoresports

A&E

Careers

Classifieds

New In Homes

Photos

Shopping

Travel

Wheels

14-day site search

[More search options](#)

Stock Quote

☐ Ticker ☐ Name

Members

[Notice to our readers](#)

Business

[@Biz](#)

[Money 301](#)

[Cribb Notes](#)

[Today's Paper](#)

[Star Columnists](#)

Business Extras

[Portfolio Manager](#)

[Today's Markets](#)

[Most Actives/Highs & Lows](#)

[Stock Quotes](#)

[Fund Quotes](#)

[Exchange Rates](#)

[Loans & Rates](#)

[Fees & Services](#)

[Tools & Calculators](#)

> ADVERTISEMENT <

Aug. 30, 2004. 08:26 AM

Getting the lead out

RACHEL ROSS
 TECHNOLOGY REPORTER

There's a transformation underway in consumer electronics, a greening of the industry led by hard-hitting legislation.

Europe wants to get the lead out. The rest of the world is coming along for the ride.

In 2006, the European Union's Restrictions of the use of certain Hazardous Substances (RoHS) legislation will kick in, banning the use of six toxins from most electronics. Many manufacturers have already eliminated some

MARKET LINKS

- > [Today's markets](#)
- > [Stock quotes](#)
- > [Mutual fund quotes](#)
- > [Portfolio manager](#)
- > [Bill Carriagan](#)
- > [Pat McKeough](#)
- > [David Olive](#)
- > [Ellen Roseman](#)

Stock List

Follow up to 15 stocks with Stocklist, thestar.com's free tracking tool.
[Click to launch](#)

Article Submission

3.2 Automatic metadata markup

The EHN taxonomy has also been tagged with matching rules, which allows the site to automatically tag the submission with dc:subject properties from the taxonomy. For the first version of the new EHN site, the rules are simply regular expressions which match against the article text. While it is planned in the future to use more sophisticated named-entity extraction techniques, this simple approach has proven effective – editors estimate that 90% of the effort involved in manual categorization is saved. The automatic categorization occurs when the contributor selects the text in the right-hand pane, and clicks "Grab Text." A third pane appears on the bottom left, displaying the article text for the editor to review. The results of the categorization appear on the left.

iswc2004.semanticweb.org/demos/07/index.html

5/10

The screenshot shows a web browser window titled "TheStar.com - Getting the lead out' - Post to EHN - Mozilla Firefox". The browser's address bar shows "Read ehndev.siderean.com".

Left Sidebar (Classification):

Add Subject Classification of type: (Search)

Select one

The following categories were assigned automatically:

- ☒ Exposure pathway > **Soil/sediment**
- ☒ Current issues > Children's health > **Cognitive/behavioral in children**
- ☒ Solutions > **Law**
- ☒ Current issues > Energy > By energy use > Transportation > **Automobile/trucks**
- ☒ Infrastructure > **Waste disposal/landfill**
- ☒ Infrastructure > Waste disposal/landfill > **Recycling**
- ☒ Human health condition > **Cognitive/behavioral**
- ☒ Current issues > Environmental politics > **Legislative**
- ☒ Human health condition > **Poisoning**
- ☒ Exposure pathway > **Intentional exposure**
- ☒ Exposure pathway > **Accidental exposure**
- ☒ Contamination agent > Metal > **Lead**
- ☒ Contamination agent > Radiation/radioactive materials > **Microwave**
- ☒ Infrastructure > Waste disposal/landfill > **Landfill**

Article access:

Open Access

Weight:

Main Content Area:

14-day site search: GO

Stock Quote: GO

More search options: ☐ Ticker ☐ Name

Members: [Previous Story](#) [Print Story](#) [E-mail Story](#) [Next Story](#)

Aug. 30, 2004. 08:26 AM

Getting the lead out

RACHEL ROSS
TECHNOLOGY REPORTER

There's a transformation underway in consumer electronics, a greening of the industry led by hard-hitting legislation.

There's a transformation underway in consumer electronics, a greening of the industry led by hard-hitting legislation.

Europe wants to get the lead out. The rest of the world is coming along for the ride.

In 2006, the European Union's Restrictions of the use of certain Hazardous Substances (RoHS) legislation will kick in, banning the use of six toxins from most electronics. Many manufacturers have already eliminated some of the toxins targeted by the restrictions, such as certain kinds of flame-retardants.

Lead, however, hasn't been so easy to replace.

After years of research and development into lead-free alternatives for electronics, the industry is finally making the transition to so-called "greener" choices.

MARKET LINKS

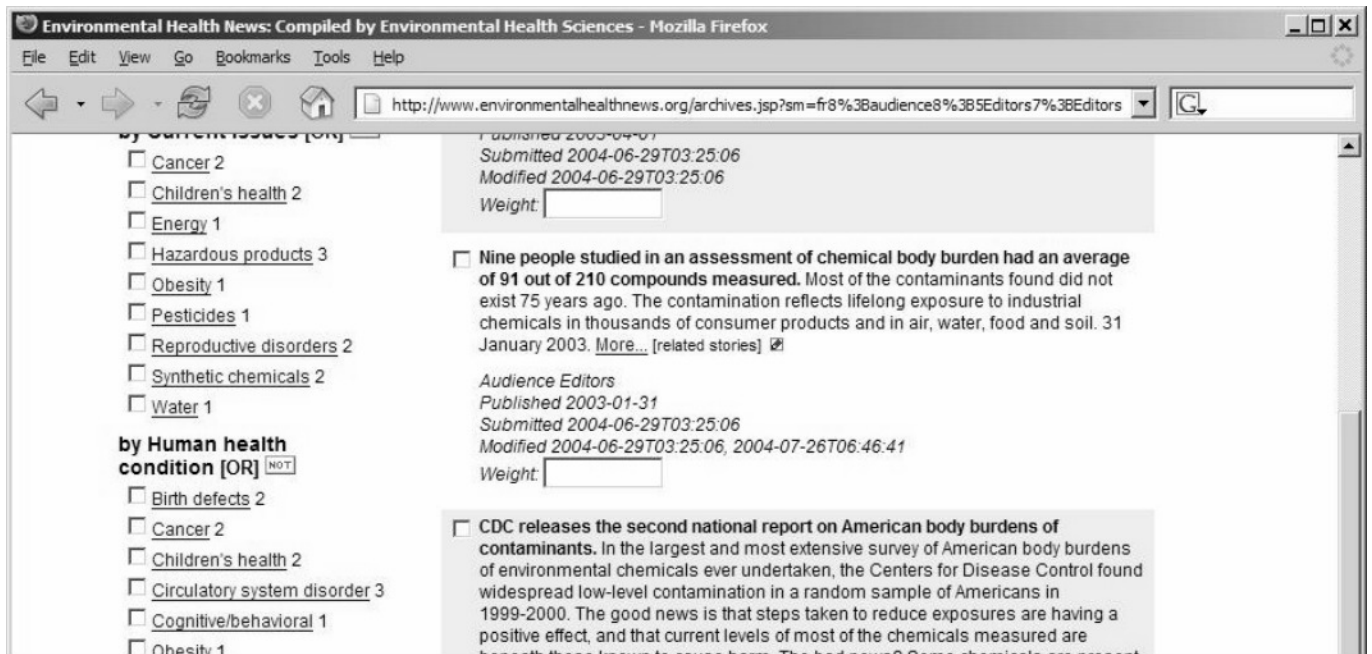
- > [Today's markets](#)
- > [Stock quotes](#)
- > [Mutual fund quotes](#)
- > [Portfolio manager](#)
- > [Bill Carriagan](#)
- > [Pat McKeough](#)

Automatic Markup

Other tasks, such as choosing related articles or assigning publisher information, are done by making use of faceted navigation pages. Typically however, the publisher is detected automatically as the article's URL is matched against a URL mask stored for the publisher.

3.3 Editorial Review

All articles submitted to the site are marked as needing editorial review and are reviewed by the editors on the archives page, which displays additional controls to the logged-in editors:



Editorial Review

This page also gives the editors the ability to mark articles for the week's top stories list or for the daily newsletter.

3.4 Weekly Top Stories

Once a number of articles have been selected for inclusion in the week's top stories list, they can be reviewed, reordered and published on the Top Stories page:

Edit Top Stories - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://ehndev.siderean.com/edittopten.jsp

Environmental Health News
Compiled by Environmental Health Sciences

Front Page Archives Subscribe Syndication About EHN Help Search Go

Preview Front Page Submissions Teasers Edit Top Stories Edit Newsletter Publishers Config Users Logout

1 31 August **Battle over toxic metal** Over the objections of several federal scientists, the Bush administration is preparing to relax national standards for selenium - a toxic metal that caused mass deformities of water fowl in California's Central Valley during the 1980s. [Sacramento Bee](#), California. ☒

2 30 August **Getting the lead out** Europe wants to get 6 toxic substances out of most electronics. The new regulation is spawning a world-wide transition across the industry. [Toronto Star](#), Ontario. ☒

3 31 August **The Man Who Knew Too Much** A BU professor fights to publicize his report claiming there's a link between cancer deaths and IBM plants. But Big Blue is doing all it can to stop him. ☒

4 31 August **The fall of a pesticide policeman** The troubled tenure of a state inspector provides a window into the state's efforts to safeguard human health and the environment as farm chemicals are sprayed in the spring and summer across Washington's farmlands. [Seattle Times](#), Washington. ☒

Top Stories for:

Done

Top Stories

3.5 Daily Newsletter

Similar to the Top Stories, the selection of articles marked for inclusion in the day's newsletter can be reviewed and reordered on the newsletter page, prior to publication. The "Teaser List," additional information summarizing the rest of the day's content on the site, is also displayed as it will be included in the newsletter. Other static newsletter text used to frame and format the newsletter is specified on a separate page.

Edit Newsletter - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://ehndev.siderean.com/editnl.jsp

Environmental Health News
Compiled by Environmental Health Sciences

Front Page Archives Subscribe Syndication About EHN Help Search Go

Preview Front Page Submissions Teasers Edit Top Stories Edit Newsletter Publishers Config Users Logout

1 30 August Getting the lead out Europe wants to get 6 toxic substances out of most electronics. The new regulation is spawning a world-wide transition across the industry. [Toronto Star](#), Ontario.

2 31 August Battle over toxic metal Over the objections of several federal scientists, the Bush administration is preparing to relax national standards for selenium - a toxic metal that caused mass deformities of water fowl in California's Central Valley during the 1980s. [Sacramento Bee](#), California.

3 31 August The fall of a pesticide policeman The troubled tenure of a state inspector provides a window into the state's efforts to safeguard human health and the environment as farm chemicals are sprayed in the spring and summer across Washington's farmlands. [Seattle Times](#), Washington.

4 31 August The Man Who Knew Too Much A BU professor fights to publicize his report claiming there's a link between cancer deaths and IBM plants. But Big Blue is doing all it can to stop him.

Today's Teaser Block
More news from today

- Gene expression and sex differences; Diabetes rampant among Native Americans
- Nanotoxicology
- Locusts in Senegal; Clean air in Japan; Huge dam in Laos
- Gold mining/Air pollution in Indonesia; Water/air in India
- Air in PA, NC, KY, IL, OR, CA, AK

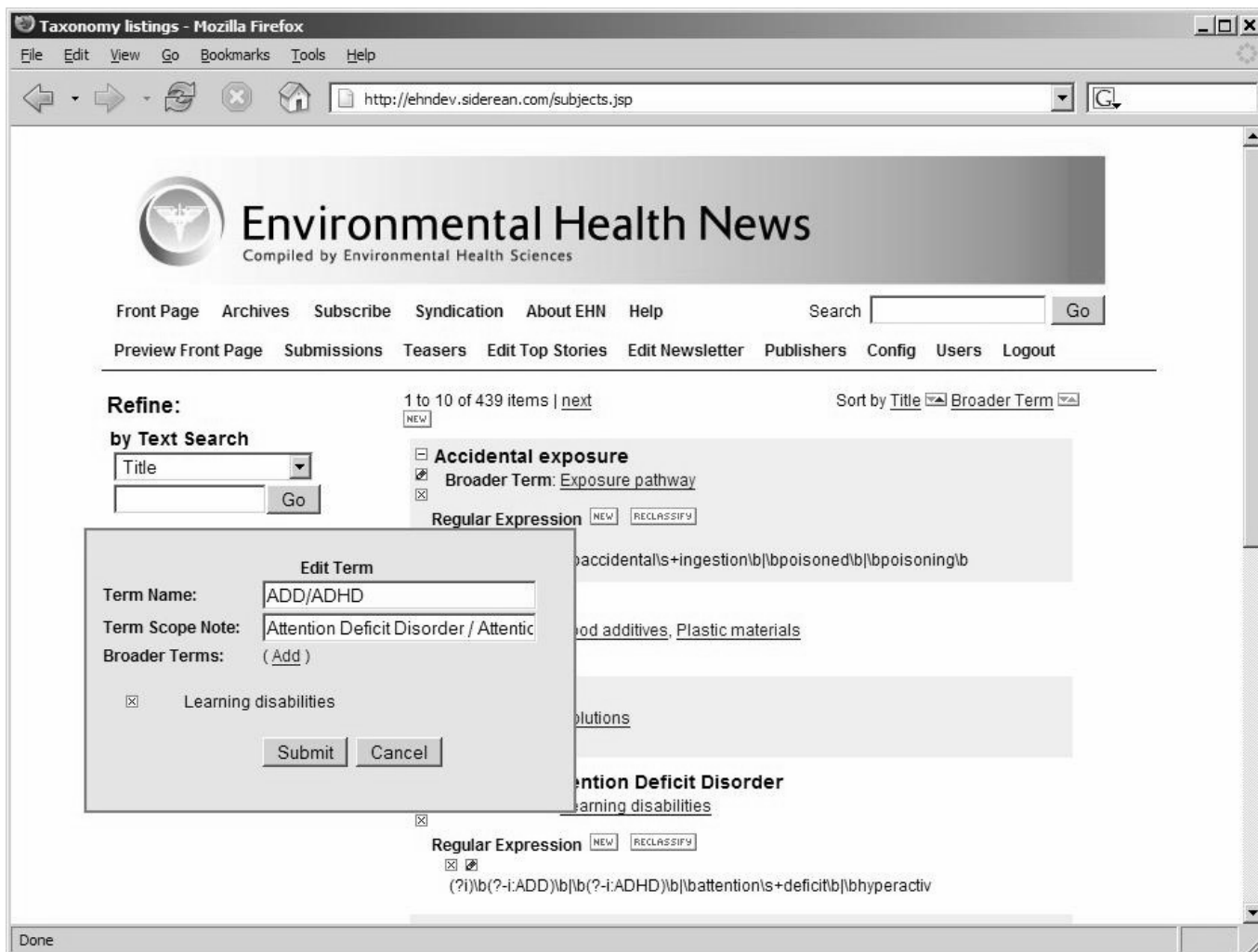
Update Drop unranked items

Email preview to: Send ☐ Plaintext only Publish newsletter

Newsletter Page

3.6 Taxonomy Editing

The latest version of the EHN site also provides the site editors with the ability to manipulate and edit the EHN taxonomy directly, rather than by external means. Using the taxonomy editing page, terms can be added, removed, renamed and re-parented. Article text-matching rules associated with the terms (which allow for automatic classification of articles as they are submitted) can also be added, edited or deleted.



Taxonomy Editing

4. Metadata

4.1 RDF Vocabulary

The metadata representing the content on the EHN site is represented using standardized namespaces where appropriate. This includes the DCMI's Element Set [2] and Metadata Terms [3] to describe basic information about each article described on the site, SWAD-Europe's TIF [4] for representing the EHN Taxonomy and the Prism element set [5], to cover properties particular to publishing, which Dublin Core does not cover. Additionally, the EHN site uses the following RDF Classes to describe its resources:

- `dcmitype:Service`
used to describe article publishers
- `ehn:Post`
used to describe individual submissions. Subclasses include `ehn:Story` for News items, `ehn:Editorial` & `ehn:Opinion` for editorials & opinion articles, `ehn:Study` for Scientific Studies and `ehn:Report` for Organization Reports
- `ehn:PostCollection`
represents a collection of posts. Subclasses include `ehn:Newsletter` which represents the Newsletters published daily by the editors, and `ehn:TopTenList` which is used to represent each week's list of top stories

5. References

- [1] Jennifer English, Marti Hearst, Rashmi Sinha, Kirsten Swearingen, Ka-Ping Yee: *Flexible Search and Navigation using Faceted Metadata*, January 2002
- [2] *Dublin Core Element Set, Version 1.1*, Dublin Core Metadata Initiative, <http://dublincore.org/documents/dces/>
- [3] *DCMI Metadata Terms*, Dublin Core Metadata Initiative, <http://dublincore.org/documents/dcmi-terms/>
- [4] *Thesaurus Interchange Format (TIF)*, SWAD-Europe, <http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/tif.html>
- [5] *Prism -- Publishing Requirements for Industry Standard Metadata*, <http://www.prismstandard.org/about/>