

Geophysical Research Abstracts,
Vol. 10, EGU2008-A-11514, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-11514
EGU General Assembly 2008
© Author(s) 2008



Harmonization and Integration of Semi-Structured Data through Controlled Tagging

E.M. Robinson and **R.B. Husar**

Washington University, St. Louis, Missouri, USA (rhusar@me.wustl.edu / Phone: +1 314-935-6099)

The contents of cyberspace are increasingly generated and distributed by individuals. This is manifested by the explosive growth of web-based social software like wikis, media-sharing services and blogs. This architectural, technological and cultural transformation of the Internet, commonly referred to as Web 2.0, is good news for the Earth Science community since it offers new possibilities for sharing and harvesting community-provided content as well as collaboratively creating new things. One key feature of all of these new softwares is the end-user's ability to add tags, adding value by extending the metadata of the particular object. Ad hoc tagging (folksonomy) gives a rich description of the internet resources, but it has the disadvantage of providing a fuzzy schema. The semantic uniformity of the internet resources can be improved by controlled tagging which apply a consistent namespace and tag combinations to diverse objects. We have used the above tagging approaches in order to gather internet resources pertaining to air quality events. Initial event analysis of the southern Georgia fires, which burned in April and May, 2007, began with filtering and harvesting user-contributed web content. The Google Blog Search of 'Florida smoke' returned several thousand entries, many of them unrelated to the wildfires. Visually scanning the blog entries yielded a number of interesting posts, which were given the controlled tags '070508+Florida+Smoke' in the social bookmarking tool del.icio.us. Additional smoke photos were found in the photo-sharing service, Flickr and given the same set of controlled tags. Together, these tools yielded a rich but only qualitative description of the Georgia Fires. Because of the common set of controlled tags these web objects (i.e. links and photos) were harvested in a wiki environment, which also contained the

links to quantitative air quality analysis based on satellite and surface observations.