

iSERVO Semantic Scholars' Grid

Geoffrey Fox
Indiana University
gcf@indiana.edu

We discuss a world full of Geoscientists and their documents and how their science will be enhanced by tools that relate people, documents and science. The resultant picture suggests the future of universities and publishers as they adapt to the power of the Internet. We introduce

- a) PubiSERVO which is the federated distributed global database containing all catalogued documents and electronic entities relevant to iSERVO. The system stores both documents and web resources with data at latter archived
- b) The PubiSERVO databases are supported by Google Scholar style crawlers that scour the web for everything of conceivable interest to iSERVO. We generalize GML to iSERVOML specifying all relevant semantic information for iSERVO. This plus the relevant non-Geoscience specific information in documents (Authors, Institutions, Titles, Citations, abstracts, keywords – roughly the Dublin Core and the publisher's PRISM <http://www.prismstandard.org/>) represents the semantics of the information in PubiSERVO
- c) The need for human curation is minimized but some will probably be necessary to guarantee quality of PubiSERVO
- d) Geosocial tagging is supported on PubiSERVO allowing Geoscientists (and their relevant colleagues) to tag (apply keywords) in the fashion of <http://www.connotea.org> and related sites (<http://del.icio.us/>, <http://www.bibsonomy.org/> and <http://www.CiteULike.org>). Groups are supported and PubiSERVO has resources to encourage the same tag name to be used broadly for a given topic
- e) The system will be open so everybody can contribute tools to analyze the information in PubiSERVO. Initial tools will support link analysis to cluster people, documents and scientific concepts stored in PubiSERVO. Portal interfaces will support appropriate visualization using the Geographical Information Systems Grid for any geospatially-labeled data.

The publishers of the future will presumably contribute documents freely to PubiSERVO. Their business model instead involves services – Refereeing, Semantic extraction tools in loading PubiSERVO, Curating PubiSERVO and high value analysis tools.

Universities will adjust to such electronic tools which continue the Internet's democratization of Science. The value of people and their documents is measured by their proximity in PubiSERVO space – not in geographic collocation.

We discuss progress with a prototype being built for ChemInformatics and NIH's PubChem resource.