Geophysical Research Abstracts
Vol. 14, EGU2012-14465, 2012
EGU General Assembly 2012

The Changing Conduct of Geoscience in a Data Intensive World (Ian McHarg Medal Lecture)

P. Fox
Tetherless World Constellation, RPI.

Electronic facilitation of scientific research (often called eResearch or eScience) is increasingly prevelant in geosciences. Among the consequences of new and diversifying means of complex (*) data generation is that as many branches of science have become data-intensive (so-called fourth paradigm), they in turn broaden their long-tail distributions - smaller volume, but often complex data, will always lead to excellent science. There are many familar informatics functions that enable the conduct of science (by specialists or non-specialists) in this new regime. For example, the need for any user to be able to discover relations among and between the results of data analyses and informational queries. Unfortunately, true science exploration, for example visual discovery, over complex data remains more of an art form than an easily conducted practice. In general, the resource costs of creating useful visualizations has been increasing. Less than 10 years ago, it was assessed that data-centric science required a rough split between the time to generate, analyze, and publish data and the science based on that data. Today however, the visualization and analysis component has become a bottleneck, requiring considerably more of the overall effort and this trend will continue. Potentially even worse, is the choice to simplify analyses to 'get the work out'. Extra effort to make data understandable, something that should be routine, is now consuming considerable resources that could be used for many other purposes. It is now time to change that trend.

This contribution lays out informatics paths for truly 'exploratory' conduct of science cast in the present and rapidly changing reality of Web/Internet-based data and software infrastructures. A logical consequence of these paths is that the people working in this new mode of research, i.e. data scientists, require additional and different education to become effective and routine users of new informatics capabilities. One goal is to achieve the same fluency that researchers may have in lab techniques, instrument utilization, model development and use, etc. Thus, in conclusion, curriculum and skill requirements for data scientists will be presented and discussed.

* complex/ intensive = large volume, multi-scale, multi-modal, multi-dimensional, multi-disciplinary, and heterogeneous structure.