

Broad Data: What happens when the Web of Data becomes real?

Jim Hendler *

* Rensselaer Polytechnic Institute, USA
hendler@cs.rpi.edu,
<http://www.cs.rpi.edu/hendler>

Summary

“Big Data” is used to refer to the very large datasets generated by scientists, to the many petabytes of data held by companies like Facebook and Google, and to analyzing real-time data assets like the stream of twitter messages emerging from events around the world. Key areas of interest include technologies to manage much larger datasets (cf. NoSQL), technologies for the visualization and analysis of databases, cloud-based data management and datamining algorithms.

Recently, however, we have begun to see the emergence of another, and equally compelling data challenge – that of the “Broad data” that emerges from millions and millions of raw datasets available on the World Wide Web. For broad data the new challenges that emerge include Web-scale data search and discovery, rapid and potentially ad hoc integration of datasets, visualization and analysis of only-partially modeled datasets, and issues relating to the policies for data use, reuse and combination. In this talk, we present the broad data challenge and discuss potential starting points for solutions. We illustrate these approaches using data from a “meta-catalog” of over 1,000,000 open datasets that have been collected from about two hundred governments from around the world.

Biography

Jim Hendler is the Tetherless World Professor of Computer and Cognitive Science, and the Director of the Institute for Data Exploration and Applications at Rensselaer Polytechnic Institute. He is also a faculty affiliate of the Experimental Multimedia Performing Arts Center (EMPAC), serves as a Director of the UK’s charitable Web Science Trust and is a visiting Professor at DeMontfort University in Leicester, UK. Hendler, a fellow of IEEE, AAAI, BCS and AAAS, is the first computer scientist to serve on the Board of Reviewing Editors for Science. He was named one of the 20 most innovative professors in America by Playboy magazine and was selected as an “Internet Web Expert” by the US government.