

New application allows scientists easy access to important government data

December 10 2010



Rensselaer collaborated with Elsevier on the application. Credit: Rensselaer Polytechnic Institute

Government agencies around the world make billions of bits of raw data available to the public each day, but this data is often in difficult formats or so widely spread around the Web it is virtually unusable to the public and scientists who seek to use this valuable information in their research.

Computer scientists within the Tetherless World Research Constellation at Rensselaer Polytechnic Institute have developed an application to help solve the problem. A collaboration with scientific publisher Elsevier, the

application utilizes the U.S. [government data](#) warehouse, Data.gov, to provide scientists with easy and direct access to government data sets relevant to their research.

For Rensselaer, the work is the latest example of the renowned Web Science research group's efforts to enhance the hundreds of thousands of raw government datasets available on the Data.gov website with advanced [Semantic Web](#) technology. Their work is bringing scientists and the public usable, relevant, searchable, and easy replicable datasets on topics from [climate change](#) to public safety to the federal deficit.

The new application, called US Government Dataset Search, lives on Elsevier's SciVerse websites. SciVerse provides the global scientific research community with searchable access to the world's largest source of peer-reviewed scientific content. Such access is a vital component of the modern scientific process as scientists develop new discoveries by building off the findings of previous peer-reviewed publications.

"There is a growing movement to make data and content more open and accessible on the Web," said Tetherless World Research Constellation Professor James Hendler. "Elsevier's tool-based systems show a new way for publishers to join this movement without sacrificing copyrights. It should serve as a starting place to be emulated by others around the world."

Once selected from an application gallery by SciVerse users, the new application will display a customized list of government data sets most relevant to the topics for which the scientist is searching for articles. As an example, a climatologist searching SciVerse for peer-reviewed articles on climate change would be provided with a list of all relevant government data on Data.gov ranging from the National Oceanic and Atmospheric Administration's massive collaborative weather observation networks to historical climate diaries and journals from the National

Archives. This free and relevant data can then be used by the scientists to advance their research, often in totally new and unexpected ways, according to its developers.

In addition to providing direct access to raw government datasets, the application simultaneously searches the Linking Open Government Data (LOGD) portal at Rensselaer's Tetherless World Research Constellation. The portal hosts Data.gov datasets that have been converted and enhanced with Semantic Web technologies. Semantic enhancements to the datasets make them much more usable and searchable to a variety of applications, enabling multiple data sets to be linked even when the underlying structure or format of each is different. Completely unseen to the average user, this semantic technology resides below the surface of the Web, augmenting rather than replacing traditional search engines. [Computer scientists](#) and developers can also take the semantic coding and utilize and enhance it independently.

"When we enhance data with semantics, we make it much more usable to a researcher than raw data," said the project lead for the application and Rensselaer research engineer John Erickson. "Through this application and others developed within the Tetherless World, we are empowering researchers with new tools for the basic practice of science by introducing semantics into the exploration of data."

Erickson was joined in the research by research scientist Li Ding, graduate student Dominic DiFranzo, as well the professors who lead the research group, Deborah McGuinness, Hendler, and Peter Fox.

"Using Semantic Web technologies, Tetherless World Research Constellation at Rensselaer has built innovative solutions leveraging open government datasets from Data.gov," said Vice President of Product Management for Elsevier's Application Marketplace and Developer Network Rafael Sidi. "We are delighted to partner with them to bring

government datasets to our users. The Dataset Search application built by Rensselaer illustrates how collaboration with the research community can lead to innovative applications that enhance scientists' productivity."

More information: For more information on the Tetherless World Research Constellation work with Data.gov, go to data.gov.tw.rpi.edu/wiki

Provided by Rensselaer Polytechnic Institute

Citation: New application allows scientists easy access to important government data (2010, December 10) retrieved 3 May 2024 from <https://phys.org/news/2010-12-application-scientists-easy-access-important.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--