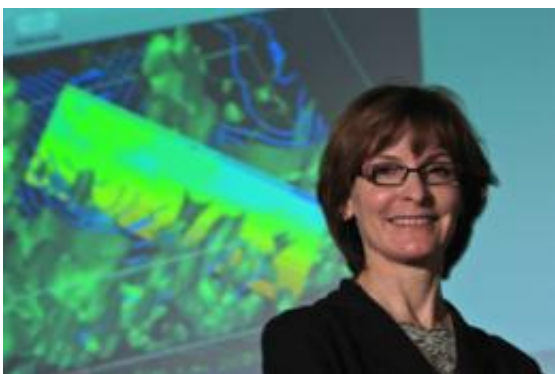


New software helps researchers find meaning in massive scientific data sets

May 24 2010



Beth Plale. Photo by Chris Meyer

(PhysOrg.com) -- One of the biggest challenges today's scientists face is sorting and making sense of the massive amounts of data produced by advanced scientific instruments and supercomputers. In response to these challenges, the IU Data to Insight Center (D2I) has released XMC Cat, a new software tool designed to make this critical task more manageable and reduce the time between data collection and possible scientific breakthrough.

XMC Cat is a catalog of metadata, or "data about data." Metadata help scientists more quickly locate the data most useful to their research. XMC Cat further accelerates this process by cataloging detailed metadata and providing access to that metadata through an easy-to-use web interface.

"For researchers, finding the right data can be a bit like looking for a scientific needle in a massive digital haystack," said Beth Plale, associate professor of computer science in the IU School of Informatics and Computing and D2I director. "XMC Cat breaks that stack into manageable, well organized sections, making it much easier for scientists to sort through and find what they need."

XMC Cat lead developer Scott Jensen noted that what makes XMC Cat so powerful is its ability to adapt to the languages used by various scientific communities, instead of requiring the user to learn a great deal of specialized knowledge.

"Many scientific communities have developed their own metadata schemas and vocabularies to describe their data," Jensen said. "XMC Cat is architected to adapt to these various schemas -- so unlike similar tools, it adapts to the scientific community, rather than requiring the community to adapt to the software. It also provides scientists point-and-click access to data without requiring them to learn new query languages or command-line tools."

Other features of XMC Cat include:

- A web-based wizard that walks the user through the process of building configuration files from a metadata schema, which then configures the catalog at installation.
- A point-and-click query interface that adapts automatically to concepts contained in the user community schema. This allows scientists to query the metadata by selecting familiar concepts and using the standard vocabulary of their scientific discipline.

- The ability to share query definitions. This is useful for locating certain model configurations or combinations of environment variables that may cause a particular model to become unstable or generate anomalous results. With XMC Cat, scientists can share their queries with others, who can in turn run it against their private data collections to see if any experiments could be impacted -- always good to know before you publish!
- Data remain private and stay in scientists' workspaces until they are made public.
- Additional metadata can be added quickly, easily, and incrementally to the existing catalog of an experiment or data set, even when a scientist is running long experiments or workflows. Metadata, as well as archived data, can be used to monitor ongoing experiments.
- A simple plug-in interface allows scientists to add modules that automatically harvest additional metadata from files -- such as experiment or workflow configuration files or the headers of binary formats such as NetCDF, HDF, or FITS.

More information: To learn more about XMC Cat, visit:
www.dataandsearch.org/dsi/xmccat

Provided by Indiana University

Citation: New software helps researchers find meaning in massive scientific data sets (2010, May 24) retrieved 25 April 2024 from <https://phys.org/news/2010-05-software-massive-scientific.html>

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.