Conlon, Mike, "The Objects of Science and Their Representation in eScience"

The objects of science and their representation in eScience

Mike Conlon
University of Florida
mconlon@ufl.edu, http://vivo.ufl.edu/individual/mconlon

Science is done by people. They form teams, make hypotheses, write grants, build and use tools, observe nature, conduct experiments, collect data, draw pictures, analyze, draw conclusions, present results, write papers, and generate data and other artifacts. They teach, mentor and model the next generation of scientists. How does eScience -- the application of information technology to scientific processes -- help, hinder, or change science? What is missing? What can we recommend for improvements?

Consider the "objects" of science as in modeling. A simple model might start with people, funding, resources, papers and data. For eScience to help, it might provide a means to know about science and to facilitate the scientific work in any of the aforementioned processes.

People

Systems such as VIVO are being created to identify people, create representations of those people, model expertise, interests and activities, and to associate those representations with other objects in science. A rich vocabulary of connections between people is envisioned (coauthoredWith, coPIWith, mentored, taughtBy). Information about people, their connections to each other and their connections to other objects in science can then be used to build teams in support of the various science processes around the world, and to better understand the nature of scientific collaboration and predict future productivity.

Funding

Simple systems are available to model funding to provide information about who got which resources to perform what work. US federal science agencies should provide this data in a common open format. Collaboration with EU and other funding agencies should provide significant world coverage. Connections can then be made between the elements of funding and it's first order products -- papers, data and tools. Connections can then also be made to second order products and beyond -- better health, improved economic or social conditions. Corporate, private and other funding will require disclosure at appropriate points in the scientific process.

Resources

Systems are being created to model research resources -- equipment, computational resources, cell lines, tools, software, core labs, national labs and more. These resources can be connected to the people who created them, manage them, and use them. Resource discovery results. The Eagle-I project and others are developing semantics for representing resources and their connections to people, funding and other objects of science. Support is needed to demonstrate the utility of such information resources in a variety of settings.

Papers and other creative works

Much work has been done to make scientific literature available in a variety of formats and contexts. Standard, open, semantically clear citation objects have been designed, but are only partially available. A world resource of open citation information is needed to form a basis for connecting (attributing) people to papers. Publishers, aggregators and governments all have an interest in making this data freely available. Simple attribution (authorship) can be augmented by a richer set of connections of people to works. A new vocabulary of attribution will enhance our understanding of the contributions made by scientists to works.

Data

Scientists create and analyze data. They can create additional findings from data previously collected. Well-curated, publicly available data is rare. Scientists need incentives to make data available to others via collaboration or more generally. Identification, semantic description, curation, citation, access control and attribution are all frontier areas for the use and reuse of data in science. Repositories are beginning to make inroads in the provision of data. Incentives potentially unlock data for additional reuse. Linking data to papers and people will provide new opportunities for collaboration and reuse.

Connections, tools and opportunities

Connecting objects to each other opens a world of new expression, resulting tools and opportunities for discovery. The efficiency of science goes up -- new collaborations are formed, additional data is made available for additional findings. Papers and works are more easily found and can be more readily assessed having been associated with data, funding, people or resources of interest. An observatory for the objects of science can be created giving us a window on science as it operates and to see the operation as it occurs. The impact of science is more readily determined when objects and their connections can be visualized and assessed.

More objects, future work

One can easily imagine additional objects to be added to an eScience framework -- hypotheses, assertions, concepts, provenance, ownership, patents, institutions, and many more. By committing to open reusable models, software and data, we can build an eScience infrastructure that accelerates discovery.

Recommendations

1. Support development of open systems for representing the objects of science, including object modeling, semantics, and their technical implementations.
2. Support institutional adoption of open systems for representing the objects of science.
3. Support development of tools and retrofitting of existing tools to use the objects of science for more efficient science and for information about science.
4. Provide an international open standard data set for bibliometric information for all published work world wide at the level of papers, possibly through a collaboration of international libraries -- Library of Congress, British Library, etc.