Universal bioinformatics system

Paula Park
Email: ppark@the-scientist.com

TORONTO - A consortium of computer companies and academics on 10 June 2002 presented a naming system for life science information that they said could simplify the identification and tracking of genes and proteins.

The Interoperable Informatics Infrastructure Consortium (I3C) demonstrated a universal nomenclature at the Biotechnology Industry Organization (BIO) 2002 Annual Convention. The Life Science Identifier (LSID) defines a simple convention for identifying and accessing biological data stored in multiple formats.

Today, researchers use more than 400 file formats, and each lab has its own system for naming and structuring the data. "[LSID] allows us to identify an object in a database or flat file and assign it a single name," said Brian Gilman, head of the I3C technical architecture working group, one of 1000 exhibitors at the convention. "We're trying to make things as open and transparent as possible."

Gilman, who is also a group leader for informatics in the medical and population genetics program at the Whitehead Institute, said I3C hopes to help gain support for a single forum for open source software and bioinformatics tools.

In January 2001, to help bioinformatics specialists focus on analyzing data rather than on simply translating it from one system to another, Sun Microsystems' Informatics Advisory Council, the Biotechnology Industry Organization (BIO), IBM, and the National Cancer Institute (NCI) created I3C. These groups loosely modeled the organization after the World Wide Web Consortium (W3C), which attempts to ensure that web pages employ universal standards and formats. I3C now consists of 75 member organizations.

I3C plans to establish common protocols by accepting and evaluating proposed standards and conventions submitted by life scientists around the world, Gilman said. On Monday, the group also demonstrated how to integrate information using the Bioinformatic Sequence Markup Language (BSML), which was developed with funding from the National Human Genome Research Institute (NHGRI).

"We are completely supportive of the idea of community-based architecture," said Caroline Kovac, general manager of IBM life science solutions group in Somers, New York. "I think that's an enabler for bioinformatics."

I3C intends to develop multiple freely available software applications that will allow researchers to build their own architecture using common principles and structures. Some bioinformatics specialists feared that if they did not develop common tools, a large company could monopolize the industry.

"Because there's a ... need for polished software tools, some big, monolithic corporation is going to suck up and squish the smaller companies," said William Van Etten, a partner in Bioteam.net, a bioinformatics consultancy based in Massachusetts. "I3C ... is an organization ... that provides a little extraction layer between me and everyone I play with that enables me to speak with them."

Van Etten praised IBM and Sun Microsystems for what he termed an "altruistic" investment in nonproprietary software. Other corporate members include Cambridge, Massachusetts-based Avaki, Incogen, of Williamsburg, Virginia, Oracle, of Redwood Shores, California, and LabBook Inc.of McLean, Virginia. Membership dues range from $1,000 for nonprofit agencies to $50,000 for sustaining memberships. "It's all about making data publicly available," Gilman said. "Why should you have to reinvent the wheel?"

References

1. I3C web site (including LSID demo), [http://www.i3c.org/]

2. BIO 2002 convention, [http://www.bio2002.org/index.asp?stay=yes]

3. Sun Microsystems' Informatics Advisory Council, [http://www.sun.com/products-n-solutions/lifesciences/docs/iac.html]

4. IBM, [http://www.ibm.com/]

5. National Cancer Institute, [http://www.nci.nih.gov/]

6. Bioinformatic Sequence Markup Language, [http://www.bsml.org/]

7. National Human Genome Research Institute, [http://www.nhgri.nih.gov/]

8. Bioteam.net, [http://bioteam.net/]

9. Avaki, [http://www.avaki.com/]

10. Incogen, [http://www.incogen.com/]

11. Oracle, [http://www.oracle.com/]

12. LabBook Inc, [http://www.labbook.com/]

This PDF file was created after publication.