Accelerating Cancer Research Using Semantics-Driven Technology

Early-phase clinical studies can lead to significant advances in cancer therapy, but the data they produce is difficult for researchers to share, especially in reusable form. A team of UK experts is developing software that makes it easy for researchers to collect and share data in a meaningful way, thereby promoting reuse of data and accelerating progress in the fight against disease.

Rapid advances in genomics hold great promise for the treatment of diseases such as cancer. A better understanding of how cells function, for instance, enables scientists to develop clinical “biomarkers” that can be used to identify cancer patients who are unlikely to respond to a particular chemotherapy. Armed with such information early on, physicians can turn to other treatment options that might prove more effective.

The pursuit of such individualized care remains a fundamental clinical challenge for cancer researchers, says Dr. James Brenton, a medical oncologist and researcher at the Cancer Research UK Cambridge Research Institute in the United Kingdom. Current chemotherapy strategies for most cancers have only limited effect, in large part because the molecular basis for drug response and resistance is not well understood.

The best opportunities to make new discoveries about drug response and resistance—and to identify biomarkers—are through the kinds of early-phase clinical studies that are being conducted by clinics and labs worldwide. But the information from these studies is hard for other researchers to obtain. Even if the data has been published electronically, its semantics—the context of collection, details about the experimental protocol, questions asked and procedures applied—usually reside in locally

**Fast Facts**

**Project Principals:**
- Prof. Jim Davies, professor of software engineering and director, Software Engineering Programme, University of Oxford
- Dr. Jeremy Gibbons, deputy director, Computing Laboratory, University of Oxford
- Dr. James D. Brenton, Functional Genomics of Ovarian Cancer Laboratory, Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

**Web Site:**
- [www.cancergrid.org/](http://www.cancergrid.org/)
- [www.cambridgecancer.org.uk/](http://www.cambridgecancer.org.uk/)
- [www.comlab.ox.ac.uk/](http://www.comlab.ox.ac.uk/)

**Profile:**
A team of computer scientists and cancer researchers from the UK’s Cancer-Grid project are developing tools that will promote effective data sharing in translational research. These tools are built on Microsoft technology, support open standards and will be made freely available to the research community.

**Microsoft External Research**

The Microsoft External Research Division within Microsoft Research partners with academia, government and industry to advance computer science, education and scientific research aimed at helping address some of the world’s most urgent and significant social and technological challenges. Along with investing cash, software, hardware and research expertise to enable ground-breaking projects worldwide, Microsoft External Research is committed to providing the advanced technologies and services needed to support every stage of the research process. Efforts are focused in four research areas—including Health and Wellbeing, which explores technologies that advance healthcare and help people make better choices about their health.

**Microsoft External Research**
http://research.microsoft.com/en-us/collaboration/
stored documentation, paper records and the memories of the researchers. This makes reusing the data—or combining it with other data to draw more powerful conclusions—prohibitively expensive. And if the semantics describing the data are incompatible, it’s nearly impossible.

When data can be combined, however, the results can be dramatic. For instance, individual studies of the breast cancer drug tamoxifen failed to demonstrate any predictable benefit. In some patients, it appeared to prevent recurrence, but in others, it didn’t seem to help at all. An international community of researchers known as the Early Breast Cancer Trialists’ Collaborative Group was able to combine data from 42 different studies, and in so doing they discovered a biomarker that could be used to predict efficacy. For the 70 percent of breast cancer sufferers who exhibit this biomarker, tamoxifen has become a part of the recommended therapy. The other 30 percent are spared ineffective treatment and possible side effects.

“Having some immediate, direct benefit for the researchers involved is essential—and automation is the key. Everyone supports the ideal of data sharing, but no one can afford to do it manually.”

Jim Davies, professor of software engineering, University of Oxford

To make it easier for cancer researchers to share their data in a meaningful way, Brenton is working with the Computing Laboratory at the University of Oxford, where the Software Engineering Group is developing semantic tools with financial and other support from Microsoft External Research. These tools help researchers capture the design of their studies as abstract models that are linked to computable representations of the data semantics. The models are then used to generate software needed to conduct the studies, including the electronic forms used to collect the data. This saves time and effort in software development and more than repays the initial investment in modeling the study protocol. It also has the benefit of making the semantics available before the study begins, allowing researchers and sponsors to coordinate study design to facilitate data sharing and reuse. Most importantly, it ensures that the semantics can be determined automatically, just by following a link from the data.

“Having some immediate, direct benefit for the researchers involved is essential,” says Jim Davies, professor of software engineering at Oxford, “and automation is the key. Everyone supports the ideal of data sharing, but no one can afford to do it manually.”

Davies’s group gained experience in developing and delivering tools to support clinical research on projects funded by the UK’s Medical Research Council, including the CancerGrid project, and a long-running collaboration with the US National Cancer Institute. Microsoft Research has been involved from the start, providing software, hardware, technical support and financial resources. The tools themselves are based on Microsoft technologies: Office SharePoint® Server 2007 for collaboration and workflow, SQL Server® 2008 for data management, ASP.NET for custom Web pages and portals, and Silverlight™ for cutting-edge Web applications.

“These are the technologies that our users want,” says Jeremy Gibbons, deputy director of the Computing Laboratory at Oxford. “They need something that works for them in their clinic, in their laboratory, in their office. They need the semantics where the studies are documented, and where the data is stored and presented. They need the semantics linked into their office documents.”

The tools developed by the Oxford team will allow users to create and reuse semantics, using familiar applications, while working to technology-independent semantic standards. Data generated using these tools can be shared with other groups that use different technologies or use the same technologies in different ways, reflecting differences in practice, culture or scientific context.

Brenton, Davies and Gibbons are working with colleagues in Oxford, Cambridge and London to demonstrate the benefits of the semantics-driven approach—and to extend the coverage and power of the semantic tools—through a range of early-phase studies on ovarian, lung and breast cancer.

Prototype versions of the tools are already being used at Cambridge for later-phase (Phase III) clinical trials. But early-phase studies represent a significantly greater challenge. “Early-phase studies work to smaller budgets,” Brenton says. “They can’t afford to spend a lot of time and money on developing software or on producing models of protocols or data. If we can give them exactly the software they need, software that will make data sharing easier, then we stand a real chance of accelerating research and saving lives.”

About the research teams:
The Cambridge Research Institute links laboratory research to the clinic. It is a collaboration between the University of Cambridge and Cancer Research UK and is funded by Hutchison Whampoa Ltd., Cambridge University, The Atlantic Philanthropies and other donors.

CancerGrid is a consortium of researchers working on open standards and software solutions for clinical cancer informatics. Led by the Universities of Oxford and Cambridge, it was funded by the Medical Research Council and the Engineering and Physical Sciences Research Council as part of the UK e-Science Programme.

© 2009 Microsoft Corporation. All rights reserved. This case study is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY. Microsoft, SharePoint, SilverLight and SQL Server are registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners.