Accelerating Cancer Research using Semantics-driven Technology
http://cancergrid.org/

University of Oxford (with Cancer Research UK)
Jim Davies, Jeremy Gibbons, James D. Brenton

Cancer is a genetic disease, and advances in genomics are revolutionizing its treatment. The identification of clinical biomarkers will help physicians determine the most effective treatment based upon a cancer’s genetic profile.

The greatest opportunity to identify biomarkers lies in early-phase experimental studies. In one such study, we found that loss of the extracellular matrix protein TGFBI causes paclitaxel resistance by decreasing stability of microtubules, helping to explain why 50% of ovarian cancer patients do not respond to this drug: to the right (E, F) we see how the expression of this protein restores sensitivity.

In a second study, we are investigating functional imaging with PET-CT as a biomarker of early resistance. To the left, we see how a radioactive tracer can be used to reveal the spread of ovarian cancer through the body: the small dark patches are secondary tumours. By measuring the response, in terms of any visible shrinkage of the tumours, to each cycle of chemotherapy, we are able to calibrate the initial response against subsequent outcomes. Images taken before and after the initial cycle may then be used to identify those patients who will not respond; these patients can then be offered other treatments.

These kind of studies involve the collection of large amounts of clinical and laboratory data, which could be used to add power to other investigations, or to inform and direct the work of other researchers. To re-use data in this way, we need detailed information about the design of the original study. With support from Microsoft External Research, we are developing tools that capture study designs as computable models. These models can be used to determine the precise meaning of the data collected and – if prepared in advance – can also be used to generate the software support required, including the electronic forms used to collect the data.

The tools themselves are based on Microsoft technologies: Office and SharePoint for collaboration and workflow, SQL Server for data management, and SilverLight for data capture and presentation. To the left we see how an InfoPath document describing the study can drive management and subsequent analysis; above we see the integration of study data semantics into Excel. We are using the studies above to demonstrate the tools, and hence the value of this semantics-driven approach in accelerating cancer research.