Data Warehousing Used to Help Find the Cause of Breast Cancer

Submitted by: Harvard Public Relations Ltd
Thursday, 25 September 2003

Windber Research Institute has chosen data warehousing technology from Teradata, a division of NCR Corporation (NYSE: NCR), to help in its research into the cause of breast and other forms of cancer. It is the first time that molecular and clinical information has been integrated in a single enterprise data warehouse for this kind of project. Windber expects to generate 50 terabytes of information every nine months.

Windber Research Institute is one of the world's most integrated, high-throughput biomedical research facilities specifically set up to study the relationship between genes, proteins and disease. The Institute is using the data from several sources, such as GenBank (DNA sequences for further analysis), PubMed (scientific literature), SWISS-PROT (protein information for further analysis), KEGG (metabolic pathways) and DIP (protein-protein interactions), which are then linked to WRI's own molecular (DNA, RNA, protein) and clinical data. These databases are all integrated into the data warehouse in order to accelerate medical research to facilitate the study of gene and protein function as related to human reproductive cancers and cardiovascular disease.

Its research is the basis for a vaccine being used in a trial to stop the recurrence and spread of breast and other forms of cancer. Plus, it has announced an emerging programme focusing on understanding the molecular mechanisms of heart disease through a multifaceted approach that is examining molecular events associated with heart disease reversal and gene and protein changes that occur during the development of asymptomatic disease in young adults.

"Biologists are learning more each day about the relationship between the genotype, the environment and disease. But finding important mechanistic relationships or associations that are clinically useful is still difficult and slow because existing databases are not seamlessly linked and are fragmentary and/or difficult to mine even by an expert. To improve the efficiency of biomedical research, we need a data management solution that permits rapid capturing, storing and analysis of data acquired from diverse sources, and this is what Teradata brings to the table," explained Dr. Somiari, chief operating officer and chief scientific officer of Windber Research Institute.

There are many genome-sequencing, protein-analysis and clinical research facilities in the United States, Europe and Japan trying to find correlations between clinical phenotypes and genotype. However, according to Dr. Somiari, "What is still missing is the seamless integration of biologically relevant information from basic research laboratories with the clinically relevant information. With Teradata we will be able to seamlessly link clinical and demographic information, DNA sequence information, protein profile, genotype, gene expression data, histopathology and radiology.

The Teradata® technology will enable Windber to store, retrieve, analyse and manage the massive amount

of data being generated at the DNA, RNA and protein levels by using large-scale biology technologies. In essence, Windber's approach will accelerate discovery and knowledgebase generation and will help bring individualised medicine to patients by identifying the patient-specific causes at the molecular level.

Dr. Somiari added, "With 30,000-35,000 genes present in humans, finding the subset of genes that are associated with the onset, progression and/or severity of a disease is challenging. We accelerate discovery by carrying out high throughput analysis of biological samples at the DNA, RNA, protein and cellular levels. We typically generate 166 MB of information from each sample, which is why we need the enterprise data warehousing solutions that Teradata provides. Windber also has a tissue repository with a capacity for 240,000 tissue samples. We expect to generate approximately 50 terabytes of data, both images and text, in 9 months. This information must be delivered on demand. Because we are continuing to enrol participants in multiple protocols, we needed data warehousing solutions that will expand with us."

Nick Jacobs, president and chief executive officer of Windber Research Institute added, "We specifically referenced and sought the same data warehousing capabilities used by Wal-Mart and the top companies worldwide. We know that Teradata is the right solution for us to keep The Windber Research Institute as a leading force in medical research. It is one more powerful element that puts us beyond the leading edge, to the bleeding edge – the place we feel we need to be to cauterise and stop breast cancer and heart disease," he said.

"We chose Teradata because it has the functionality that makes data integration from multiple sources easy. This will enable us to examine multiple variables (clinical and biological) at the same time. It supports our main goal of rapidly elucidating the underlying molecular mechanisms and find cures for cancer," according to Somiari.

According to Martin H. Hill, director of health sciences business development for the Teradata division of NCR Government Systems Corporation, "The challenges of centralising huge amounts of information into a single data warehouse was obviously a task Teradata has handled often, but, when we can contribute to making a difference that may save lives, it is truly amazing and an honour. What Windber Research Institute is doing is astonishing. We are thrilled to be a part of it, and this is just the beginning," he said.

The Windber Research Institute is already working in collaboration with the clinical and basic scientists at Walter Reed Army Medical Center, The University of Pennsylvania, Creighton University, The University of Pittsburgh and the Immunology Research Center at the Uniformed Services University of the Health Sciences in Bethesda, Maryland.

The work with Windber Research Institute is another example of Teradata's work in life sciences to help understand the root causes of disease. Teradata began working with Information Management Consultants (IMC) in 2002 to enable researchers and scientists to exponentially accelerate the pace of genetic research on mice brains that may lead to the understanding of many human conditions, including brain diseases and forms of cancer.

"Instead of taking a year to study one gene, data mining enables us to study the potential interactions of 13,000 mice genes in just one week," said Dr. Carrolee Barlow, scientist and adjunct faculty member at Salk Institute. "From this course of study, we hope to more quickly learn how to treat certain diseases in people."

High Res Images - http://www.harvardpr.com/home/article_details.asp?id=6367

More information – Simon Rothwell/Ben Griffiths – 0208 759 0005 – simonr@harvard.co.uk

About Windber Research Institute

Windber Research Institute (www.wriwindber.org) is an integrated research facility that has the unique ability to simultaneously examine the function of many genes and proteins related to reproductive cancers and heart disease. The Institute is a key component of a multi-institutional coalition consisting of the Clinical Breast Care Project at Walter Reed Army Medical Center, the Joyce Murtha Breast Care Center at the Windber Medical Center, and the Immunology Research Center at the Uniformed Services University of the Health Sciences in Bethesda, Maryland.

About Teradata Division

Teradata, a division of NCR Corporation (NYSE: NCR), is the global leader in enterprise data warehousing and enterprise analytic technologies and services. For more information, visit www.teradata.com.

About NCR Corporation

NCR Corporation (NYSE: NCR) is a leading global technology company helping businesses build stronger relationships with their customers. NCR's ATMs, retail systems, Teradata® data warehouses and IT services provide Relationship Technology™ solutions that maximise the value of customer interactions. Based in Dayton, Ohio, NCR (www.ncr.com) employs approximately 29,500 people worldwide.