

With our flagship *DataInFormation<sup>SM</sup>* suite of solutions for Image Labeling, Data Annotation, NLP Calibration and Form Data Capture, Liberty Source is proud to serve a wide range of clients across many industry verticals. Our *Solution Briefs* provide a snapshot of the challenges clients bring to us, and the outcomes we achieve for them.

## Gene Microarray and Image Analysis

Industry <b>Healthcare / Cancer Care &amp; Research</b>	Data Type <b>Flat files, DICOM Images</b>
Project Duration <b>6 Months</b>	Ongoing? <b>No</b>

### Challenge

As part of a Comparative Effectiveness Research grant project, the Principal Investigator was faced with integrating huge volumes of genetic and image data with standard patient population, diagnosis and treatment information. Microarray gene expression experiments can generate in excess of 100,000 data points per sample. while thin slice (less than 3mm) Computerized Axial Tomography (CAT) scans produce a large number of very large (image) data files. Maintaining and mining these huge datasets becomes time-consuming and costly very quickly..

### Solution

By applying data normalization methods to architect and store only unique identifiers associated with small unique record “keys”, both storage costs and processing time is reduced immensely. For example, normalizing the microarray gene expression data names resulted in a 63% reduction in storage space and provided look-up query results against a 1 billion row query in less than 3 seconds using a single, on-premise server. CAT scan image series were provided in the Digital Imaging and Communications in Medicine (“DICOM®”) format. By applying data normalization methods to metadata attached to each image “slice” (only storing what was different from the previous image), the storage requirements were significantly reduced. In addition, generating a 3-D wire frame representation of the structure of interest (in this case cancerous tumors) not only saved space, but also facilitated algorithmic size, shape and volume metrics for pre- and post-treatment comparison.

### Outcome

The project was successful in demonstrating cost saving through storage space reduction, query and analysis performance improvements and automated tumor measurement techniques.

### Gene Microarray and Image Analysis

