Using a High Content Analysis Database as a Repository for Data

Brent Samson, Bryan P. Johnston, Chandrasekaran Vasudevan, and Audra Ziegenfuss
Thermo Fisher Scientific • Pittsburgh, PA USA

Key Words

High Content Analysis

High Content Screening

Cellomics

Open Gateway

Database Management

Plate Reader

Varioskan

XML

Introduction

As the drug discovery process becomes more horizontally integrated across different functional groups in a pharmaceutical or biotechnology company, it becomes essential that these different groups have easy access to the discovery data generated by the various groups. One way to achieve this goal is to have a common repository for all of the data in the form of a database that can be queried and easily accessed. A simple software application, such as the Thermo Scientific™ OpenGateway™ Application Software, which imports external, converted data into a database that uses Thermo Scientific™ Store™ Image and Database Management Software, would be useful for those who wish to automate the process of moving such data.

Overview

Increasing visibility and access to discovery data across an organization decreases bottlenecks from data transfer between groups that could result in increased productivity. This becomes especially important when data is not just numbers, but also includes images from such image acquiring devices such as microscopes, high-content readers, gel-imagers and plate readers. The numbers and images together make meaningful data and validate each other.

Although, it is easy on paper to create a single repository of all data including images from various types of instruments, in reality it can be an overwhelming technical challenge for the personnel involved in these tasks. The most common reason is the disparate nature of the instrument types and the proprietary data formats generated by these instruments. The most common solution is to manually move images and data out of the various commercially available database solutions that function as silos and onto a server or directory that lacks the ability to be queried in a meaningful way. This can be problematic when attempting to find a specific data set or image when large numbers of data sets and images exist in that directory or server resulting in non-productive time searching for existing data.

Taking Advantage of Standard File Formats

If users of these various technologies are able to overcome the technical challenges, the advantages of such a combined database are numerous, providing easy and secure access to data that can be queried in multiple ways as and when needed.

One of the ways to overcome the challenges is to take advantage of the numerous standard formats (e.g., csv, binary, txt, tiff, jpg) in which data and images from these instruments can be output. To get data into a database that utilizes Store Image and Database Management Software, the data must first be formatted into valid XML format according to the OpenGateway Application Software schema definition.

Thermo
SCIENTIFIC

The following discussion shows how data from a multi-mode plate reader can be exported and stored in a database that is normally used for storage and retrieval of data and images from a high content analysis reader.

Single Site Storage – Plate Reader and High Content Data

High content readers are often used in parallel with plate readers with the intention of gaining a comprehensive understanding of the biology of focus. Keeping both sets of data in the same database for easy querying and image/data retrieval provides rapid access for further data validation.

Using a Thermo Scientfic™ VarioSkan™ Flash Multimode Reader to scan a 96-well plate (or other plate formats) of U2OS cells labeled with the Premo™ FUCCI Cell Cycle Sensor and treated with six different compounds, cell viability was measured using the Thermo Scientific™ SkanIt™ Software. Data was saved in the native SkanIt Software file format.

An identically treated plate was also scanned using a Thermo Scientific™ ArrayScan™ XTI High Content Analysis (HCA) Reader. Images and data were stored using Store Image and Database Management Software.

The goal was to get both sets of data into the same database managed by Store Image and Database Management Software to allow a user to access each type of data as needed from a common location and user interface.

Three essential steps are core to automatically getting the data into your database either manually or programmatically:

1. Export the VarioSkan Flash Multimode Reader data into a tab-delimited text document containing all necessary data.

2. Convert the exported text data into XML format using the GatewayExtract sample application and the OpenGateway Application Software user guide as a reference.

3. Copy the XML file to the "CellData" directory on a computer that has OpenGateway Applicaiton Software properly configured.

Saving VarioSkan Flash Multimode Reader Data for Downstream Conversion to XML Format

The first step in the process is using the SkanIt Software to save the data to a format that can be easily read by an external application. Create a report of your data containing all possible report items (Figure 1). Save it as a tab-delimited text file, Sample.txt. This is important because the different report items may contain information that is needed for inclusion into Store Image and Database Management Software, not to mention storing the data in a single file makes it easier for a sample application to read the data. The report was then saved as a text file, Sample.txt.

Figure 1. Fluorometric scan report generation with all possible report items selected.
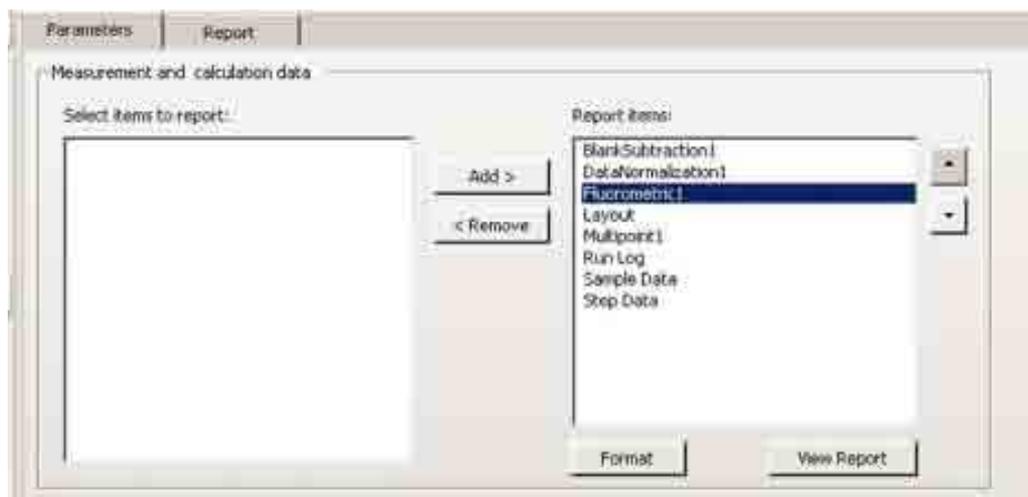
Converting VarioSkan Flash Multimode Reader Data to XML Format

The OpenGateway Software user guide contains an appendix on Data Mapping that describes the XML format. Given this XML format specification and after examining the contents of the SkanIt Software report, a sample application called GatewayExtract was created to convert the SkanIt Software text into valid XML format. For implementation details, please refer to the GatewayExtract source code. The sample application can be run using the command:

GatewayExtract.exe <inputFilename.txt> <outputFilename.xml>

ex. GatewayExtract.exe C:\VarioSkanData\sample.txt C:\CellData\sample.xml

The resulting output is an XML file that can be spooled into a database using Store Image and Database Management Software. GatewayExtract sample application code is provided for reference, as well as the executable GatewayExtract.exe, and the sample input and output. Sample code, sample data, and a sample application are all located on the instrument version of Thermo Scientific™ HCS Studio™ 2.0 (or later) Cell Analysis Software.

Spooling XML Data to a Database Using Store Image and Database Management Software

Prior to spooling data using Store Image and Database Management Software, OpenGateway Software must be properly configured. For instructions on how to do this, please refer to the OpenGateway Software user guide. Once the configuration is complete, the resulting XML file simply needs to be placed in the Thermo Scientific™ CellData directory. The file can be copied manually, or it can be written programmatically as in the previous example provided that the CellData directory is C:\CellData.

Once the data has been spooled (Figure 2), the data may be viewed using the Thermo Scientific™ View™ Application (Figure 3) similar to any other data that has been spooled using the Store Image and Database Management Software.

Figure 2. Screenshot of the View Application with spooled VarioSkan Flash Multimode Reader data appearing in the plate list (A) indicating that the data has indeed spooled to the database and is ready to be visualized.

Figure 3. Data visualization in the View Application for Feature1 demonstrating successful importation and visualization of plate reader data within an HCA database.
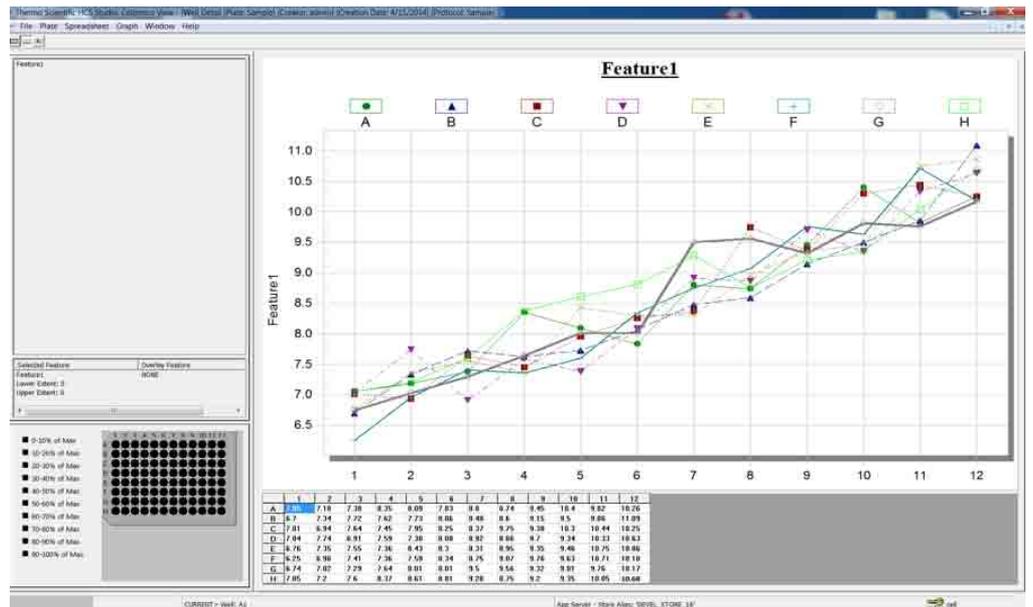
Conclusions

Storing two sets of data, such as the VarioSkan Flash Multimode Reader data and images from a high content reader, into the same database managed by Store Image and Database Management Software allows a user to access each type of data as needed from a common location and user interface. The ability to store two sets of meaningful data has the advantage of providing easy and secure access to data that can be queried in multiple ways as and when needed across an organization. By utilizing standard file formats and the XML format, OpenGateway Software can be used to overcome some of the technical challenges faced by those who wish to have a single repository for data and images. Using familiar interfaces to visualize data from multiple sources and to access this data between groups, has the potential to enhance productivity and collaborations within an organization.

thermoscientific.com/highcontent

USA +1 800 432 4091 info.cellomics@thermofisher.com

Asia +81 3 5826 1659 info.cellomics.asia@thermofisher.com

Europe +32 (0)53 85 71 84 info.cellomics.eu@thermofisher.com

C-WP_HCADR0514

Thermo
SCIENTIFIC

A Thermo Fisher Scientific Brand