Millennium\textsuperscript{32} Software and Unique Identifiers

Abstract

Millennium\textsuperscript{32} software is built using an Oracle relational database as its central data repository. This centralized database provides assistance with compliance (GLP/GMP/ER/ES)\textsuperscript{1}. All method information and raw data is stored with results. The information is individually tagged with a unique ID. Nothing is overwritten and all versions are available for review. Specific features have been implemented to allow compliance with ER/ES (21 CFR Part 11) for customers who must comply with FDA regulations.

In order to uniquely identify a result, and thus to collate all the relevant meta data, you must have:

- The correct result ID, and
- The correct project name.

These two pieces of information together allow unambiguous inspection of all methods, calibration curves and sample information used in the formation of calculated values.

Millennium\textsuperscript{32} Database

Millennium\textsuperscript{32} software is built using an Oracle relational database, called the Millennium\textsuperscript{32} database, as its central data repository. All information from projects (excluding the raw data\textsuperscript{2}), such as processed results, sample identifiers, and methods, is stored as individual data, each tagged with a unique identifier, in the Millennium\textsuperscript{32} database.

Since the individual data in the database are unique and also cannot be overwritten, all the objects stored in the database can be linked to other objects, and those connections are consistent and permanent. Consequently, Millennium\textsuperscript{32} software can reliably link raw data from chromatography instruments to the acquisition method, processing method, report method, and processed results. Conversely, Waters’ software also uses such linkages to relate results to raw data, methods, calibration curves, etc. When data are sorted in a relational database, the display of data changes in strict accordance with the relational linkages. This attribute is essential from a GLP/GMP and Part 11 compliance perspective and is also very useful in determining whether correct procedures were followed to generate a given result.

The Millennium\textsuperscript{32} database is sometimes described as \textit{channel centric}, which means that the electronic record focuses on the collection of a channel of 2D chromatographic or 3D (PDA or MS) data. Method objects must be created first before an injection can occur. In order to have a “place” for the new data to be stored in the database, Sample, Vial, and Sample Set objects are created prior to injection. At the time of sample injection, Injection and Channel objects are created and linked to the Sample objects. Concurrently, correct linkages are established to the methods used in acquisition.\textsuperscript{3}

---

\textsuperscript{1} GLP is Good Laboratory Practice, a management system to ensure the quality, integrity and validity of data underlying non-clinical laboratory studies. GMP is Good Manufacturing Practice, a quality management system to ensure product quality, safety and efficacy. “Part 11” here refers to FDA guidelines given in 21 CFR Part 11 governing electronic records (ER), and electronic signatures (ES).

\textsuperscript{2} In other words, the database contains meta-data (the project minus raw data). Project raw data is stored directly on the hard drive by default. However, the raw data path for the project can be user-specified through Configuration Manager. No user is able to modify or delete raw data except by using the Waters Service, and then only if authorized and their actions can be audit trailed.

\textsuperscript{3} If the system is in buffering mode, the linkages to the Sample Objects will be delayed.
Unique Identifiers for Millennium\textsuperscript{32} Objects

Millennium\textsuperscript{32} creates each project as a specific type of object in Oracle called an Oracle User. Objects stored in a given project are assigned IDs beginning with 1000 and incremented by one for each new object sequentially, regardless of object type. In order to keep the numerical values of Oracle identifiers manageable (a human might have difficulty working with extremely large numbers to cross-examine data, even if the database would not), a fresh set of Oracle IDs (starting at 1000) is assigned to each Millennium\textsuperscript{32} project. So, the object with ID of 1047, for example, might be an Instrument Method in one project, while in another project the very same ID of 1047 could be assigned to a Result. If data are copied between two projects, IDs are reassigned.

In order to uniquely identify a result, and thus to collate all the relevant meta data, you must have the correct result ID and the correct project name. These two pieces of information together allow unambiguous inspection of all methods, calibration curves and sample information used in the formation of calculated values.

Once the raw data is saved, the analyst may begin processing it. For example, if the analyst integrates the data, the Millennium\textsuperscript{32} software creates a simple result object. While also retaining existing links to the channel data (i.e., to the acquisition meta data), the database now adds new links to all the peak information (start, stop, retention time, height, area, integration type, etc.) and the methods used to integrate the data. In the case of a more sophisticated processing step, peak identification information, calibration curves and custom calculations may be created in the database.

If raw data are reprocessed, a new result is generated and the correct meta data are linked to the new result. New results never overwrite previous results and old results are not deleted by the system (though, it is possible for an authorized user to delete results). Each result version (for reprocessed data) is stored complete and retains all links to the other objects used in its creation. There is a counter associated with each channel that stores the number of results obtained from that channel. Each result stores the fact that it is version $x$ out of a total of $y$ versions. Thus, from any identified result, all the related meta data about how the information was collected and processed can be recalled using the relational nature of the Oracle database (see Figure 1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Waters’ software uses a relational database to relate results to raw data, methods, calibration curves and standards, sample sets, users, the actual system used to analyze the sample, etc.}
\end{figure}

\footnote{$x$ and $y$ are positive integers and $x$ is less than or equal to $y$.}