StatSoft®

Business White Paper

The *STATISTICA Enterprise-wide SPC System* (*SEWSS*): The Analysis Platform for Data Collection and Storage Systems (Part 1 of 3: Database Integration)

U.S. Headquarters: StatSoft, Inc. • 2300 E. 14th St. • Tulsa, OK 74104 • USA • (918) 749-1119 • Fax: (918) 749-2217 • info@statsoft.com • www.statsoft.com

Australia: StatSoft Pacific Pty Ltd.
Brazil: StatSoft South America
Czech Republic: StatSoft Czech Rep. s.r.o.
France: StatSoft France
Germany: StatSoft GmbH
Hungary: StatSoft Hungary Ltd.
Israel: StatSoft Israel Ltd.
Italy: StatSoft Italia srl
Japan: StatSoft Japan Inc.
Korea: StatSoft Korea
Netherlands: StatSoft Benelux BV
Poland: StatSoft Polska Sp. z o. o.
Portugal: StatSoft Iberica Ltda.
Russia: StatSoft Russia
Singapore: StatSoft Singapore
S. Africa: StatSoft S. Africa (Pty) Ltd.
Spain: StatSoft Espana
Sweden: StatSoft Scandinavia AB
Taiwan: StatSoft Taiwan
UK: StatSoft Ltd.

Table of Contents

Executive Summary

StatSoft, provider of the ***STATISTICA*** product suite, is committed to partnering with our customers in meeting our mutual goal of the design and production of products of the highest quality and reliability. The ***STATISTICA Enterprise-wide SPC System*** (***SEWSS***) is the ultimate analytics platform for research and development (R&D) and quality control. In collaboration with our customers, we learned that the major barriers to the use of analytics are:

- **Ease of use:** Regardless of the power of an analytics platform, if it is not intuitive, the vast majority of users within an organization will not adopt it.
- **Integration with Data Repositories:** Access to data often involves a time-consuming collaboration between the individuals who know the definition of the data they need and those individuals who understand the data repository and the approach to extracting the required data set.
- **Collaboration:** Research and Quality Control activities within an organization are rarely a single individual operating in isolation. Instead, these activities are programs that require collaboration among many individuals.

With this in mind, StatSoft has developed and refined ***SEWSS***, first released in 1997, over many years and iterations.

SEWSS provides a comprehensive suite of data mining, analysis and visualization tools all within a single software platform. This platform provides role-based user interfaces, analysis templates for the automation of standard analyses and reports, and collaboration through data and results-sharing within a secure environment. Every worker, technician, analyst needs access to data and analytical tools. ***SEWSS*** provides the platform to service those needs.

This paper is the first in a series of three papers that provide an introduction to the major features of ***SEWSS***, specifically focusing on its capabilities for integrating with your existing data repositories. The benefits of these features are that your organization is now provided with the necessary tools to make full and continual use of the valuable data you are collecting.

Purpose of *SEWSS*

StatSoft understands that for many organizations the status quo is not good enough in terms of taking best advantage of the wealth of data being collected and stored.

There is an acknowledgement that greater integration of analytical techniques is needed, but there is a fear that it would be expensive to develop and implement such a system and its intended users may lack the necessary training to be able to make the most use of it.

Many organizations have a wealth of existing data. Personnel within these organizations would benefit by utilizing these data, but the barriers and time commitment to do so prevent it. What is needed is an integrated system of data querying, data mining, data analysis and data visualization techniques all within a secure, multi-user, collaborative environment.

Imagine a software system where an individual simply a) has the requisite permissions and b) clicks on an analysis listed in a folder structure to be presented with the results based on the latest data. She doesn't need to know where the data are stored. She doesn't need to know that the data are being aggregated from two repositories. She doesn't need to know that the product specifications were queried separately once she selected the product and test of interest. All that she needed to know was the name of the analysis that she wanted to conduct.

That is exactly how SEWSS works!

Conceptual Introduction

SEWSS is a multi-user software application that provides a rich palette of features for use within your organization. At the highest level, the following are the major functional pieces of the ***SEWSS*** system:

- **Connections to Data Repositories**: ***SEWSS*** is configured within your organization to recognize the relevant data repositories that store the data of interest for mining, analysis and visualization.
- **Analysis Templates**: ***SEWSS*** stores definitions of analysis templates that can be manually or automatically initiated. These templates are either "hardcoded" with all of the analysis parameters specified in the template, or can be "parameterized" to allow the user of the template to select which data and specific analyses are of interest.
- **Interactive Filtering**: ***SEWSS*** provides a robust set of filtering capabilities, to hide the user from the complexities of the data structures and Structured Query Language (SQL). Instead, the user is presented with an easy-to-use menu of values (e.g., product name, Date/time range) from which she may select to determine the data of immediate interest.

- **Reports**: ***SEWSS*** provides customizable report templates to define the page layout and format for reports to be produced on a regular basis. ***SEWSS*** includes a scheduling application to manage the automation of the production of these reports at specified time intervals.
- **Security and Permissions**: In a multi-user software application, treating everyone the same is not appropriate. Roles and responsibilities within an organization differ from individual to individual. ***SEWSS*** stores account information about what the user should be able to view, modify and perform while utilizing the system. One major contributor to the system's ease-of-use is that users do not see features and reports that are irrelevant to them.

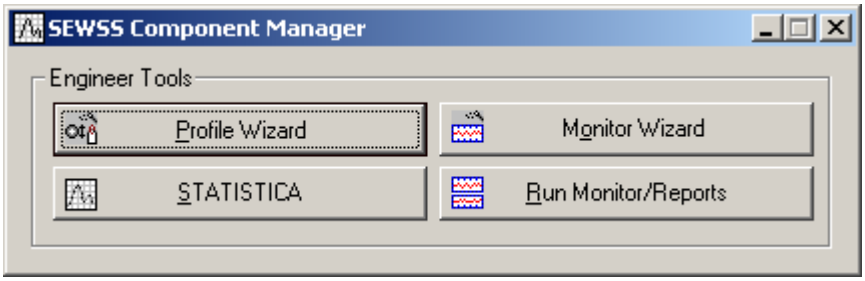
The ***SEWSS*** system provides the analysis companion to your data repositories, utilizing the above categories of capabilities. As an off-the-shelf software system, ***SEWSS*** provides general objects that are configured through its administration user interfaces. The major ***SEWSS*** objects are introduced immediately below. An understanding of the terminology used in ***SEWSS*** is an important foundation for the example that follows in the next section.

- **Connections**: Connections store the definition of a particular data repository and the method by which data are accessed.
- **Profiles**: Profiles are collections of one or more queries that define the data to be analyzed. Profiles also contain metadata for storing the necessary properties of the data to allow ***SEWSS*** to treat them appropriately for analytic purposes (e.g., that a particular column contains data, another contains the Product Name to be used for filtering, etc.)
- **Monitors**: Monitors are analysis templates that define the parameters of the analyses to be performed, including which analytical techniques to run and their respective settings.
- ***STATISTICA Visual Basic***: ***SEWSS*** is implemented as an object-oriented application with an Application Programming Interface (API) provided in Visual Basic for Applications (VBA). ***SEWSS*** customizations are available through the 11,000+ functions in the ***STATISTICA*** object model.

Example: Database Integration Capabilities

This section provides an example of the use of ***SEWSS*** as an analysis platform connecting with your organization's data repositories. In this example, we will configure a connection to a data repository, a Profile for defining the information to be analyzed, and a Monitor for defining one set of analyses to run.

Users of ***SEWSS*** must log-in and what they can view and perform is determined by the permissions of their account.

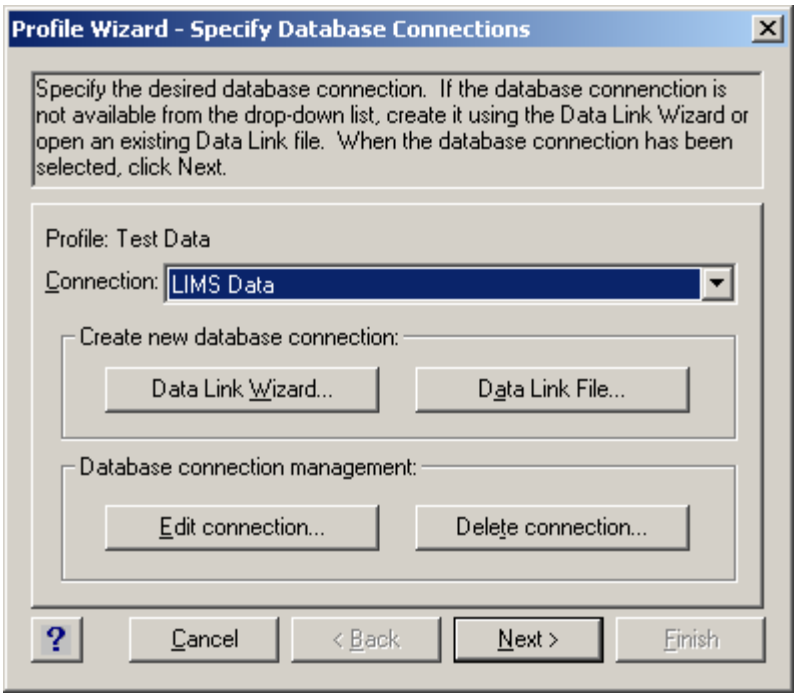


Once a user is authenticated, she is presented with the central ***SEWSS*** interface called the Component Manager.

The Component Manager provides a set of applications and utilities for analyzing data. The particular set of tools in the Component Manager depends upon the user's permissions. In this case, the user has permissions to configure Profiles and Monitors. There are Wizards provided for those configuration operations.

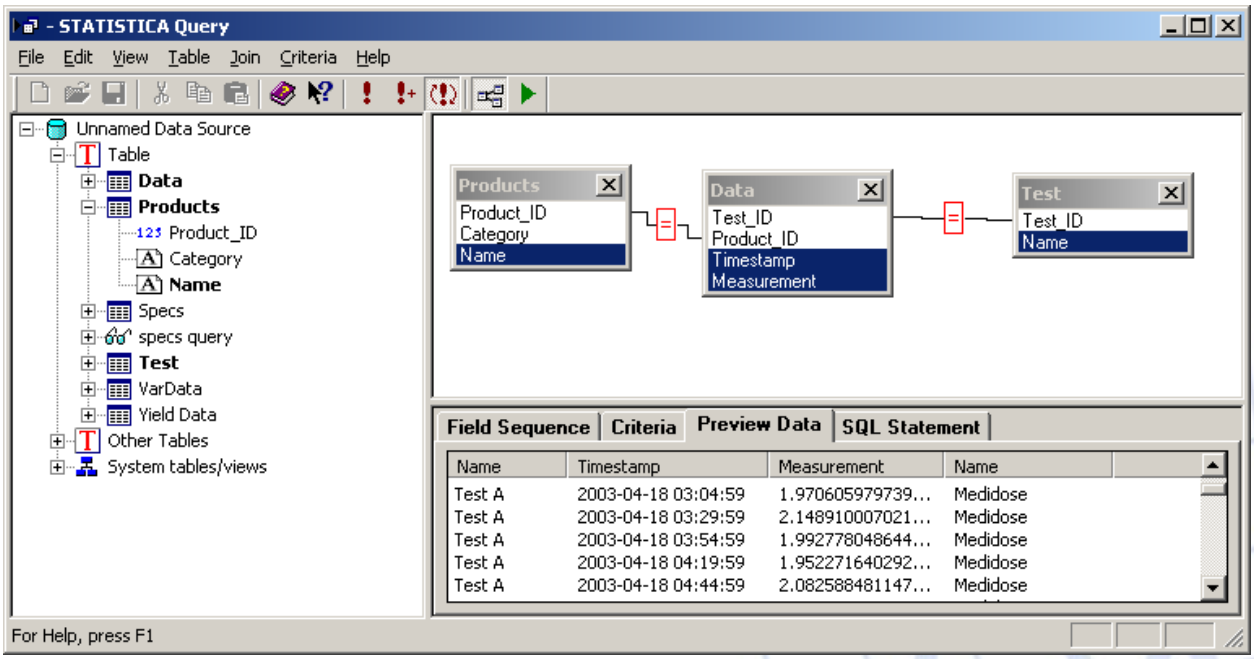
Please keep in mind that most ***SEWSS*** users are consumers of information, enjoying the pre-defined queries, analysis templates, and reports provided. ***SEWSS*** configuration is performed by a small number of system users and is performed only when a new analysis is to be setup or some similar need.

This example assumes that we have just installed the ***SEWSS*** system and are ready to connect it to our Laboratory Information Management System (LIMS) database. We begin by defining the Connection to that system.



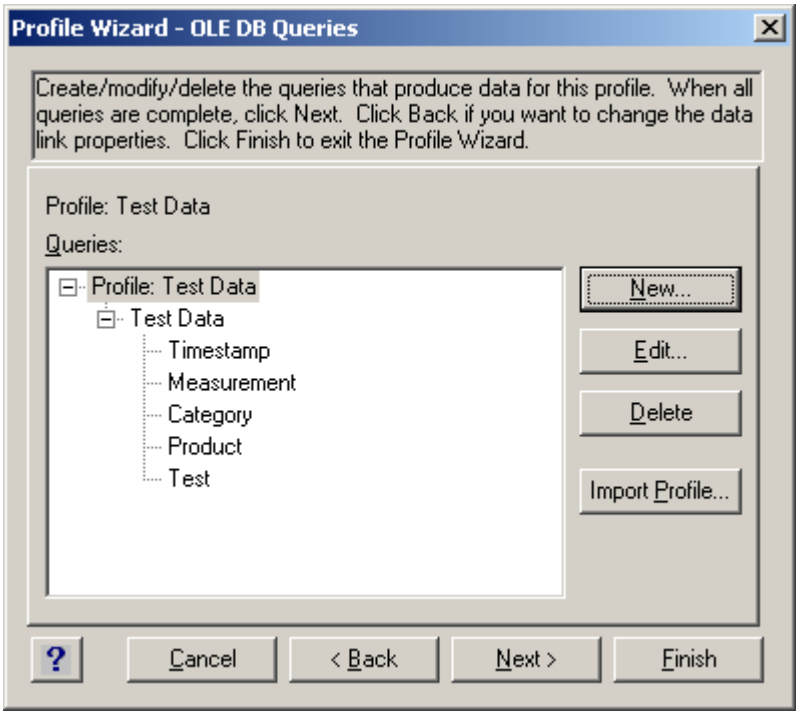
Once that Connection is defined, it can be utilized by one or more Profiles for querying data. The list of available Connections is provided to system administrators in the Connection drop-down selection box.

A Profile is a collection of Queries to define a set of data to be analyzed. ***SEWSS*** provides a user-friendly graphical query-building tool to ease the creation of these Queries, as shown below:



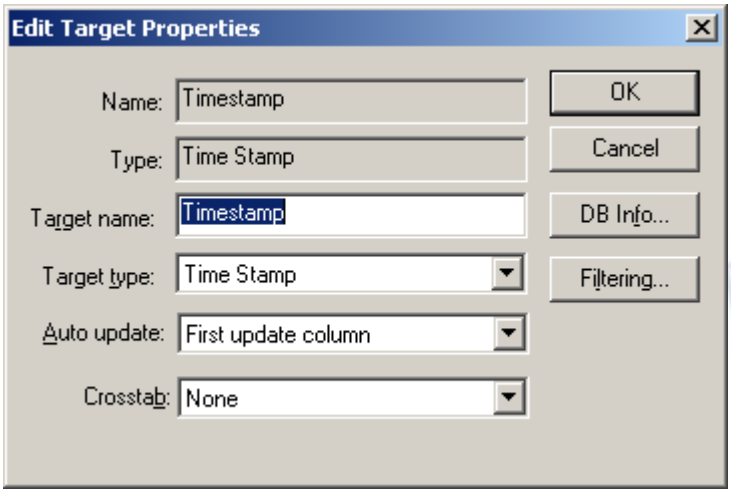
During the configuration process, ***SEWSS*** presents a Data Preview to confirm that the query results will be as expected for system users.

The set of Queries available within a Profile are displayed once they are defined within the query-building interface.

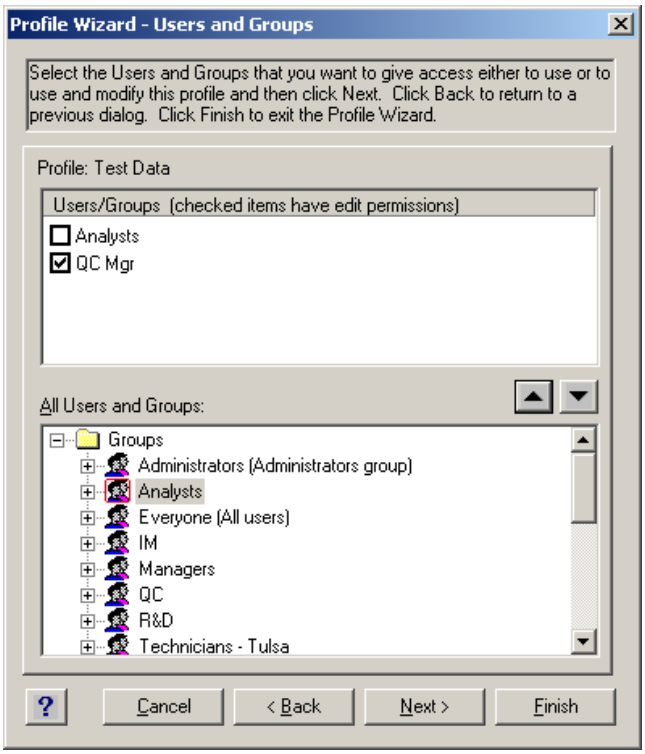


In the instance to the left, we have defined a single query called "Test Data" that returns the necessary data for a set of product tests, including the Product Name, the Test Name, the Timestamp and the Measurement values.

At this point, the administrator has the opportunity to configure the Query's metadata. This metadata layer includes the definition of how to treat the results of the query.

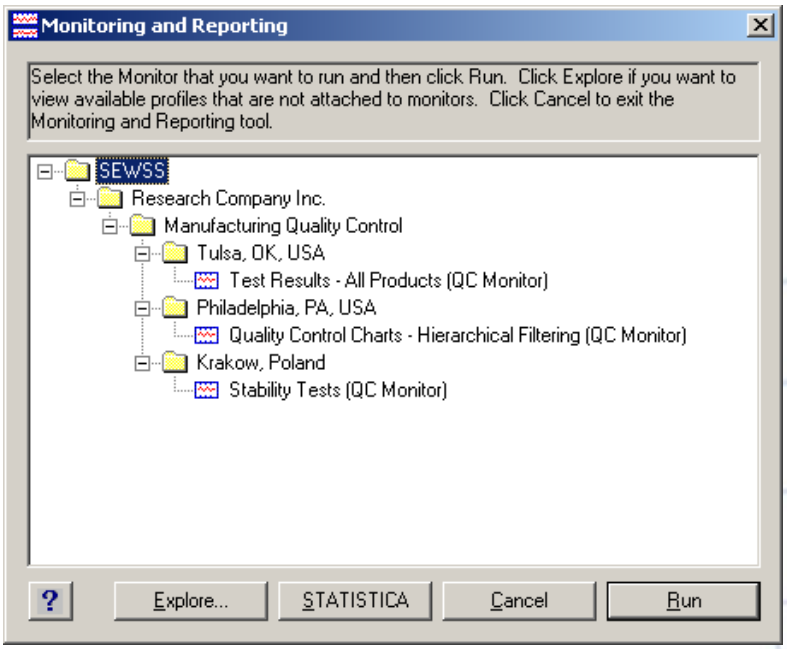


For example, to the left we have identified within ***SEWSS*** that the column called "Timestamp" provides the Time Stamps and should be used as one of the defining characteristics when running automatically updating analyses, for determining new data versus those that have already been included in the analysis.



Access to data in ***SEWSS*** is restricted based upon the user's account. When the administrator defines the Profile, she configures who should be able to utilize it.

Once a Profile is defined and stored within ***SEWSS***, one or more analysis templates may be based upon it. In other words, one of the powerful features of the system is that many analyses can be defined based upon the same data. That approach is fairly common, as an organization may wish different groups to have different views of the same data depending upon their role and interests.



Once a Monitor is defined to specify the analysis to run, consumers of analyses, who have been assigned permissions to access it, see the Monitor in the ***SEWSS*** System View.

Conclusions

SEWSS provides a powerful, flexible and easy-to-use software platform for the data mining, analytics and visualization needs of your organization. This paper provided an introduction to the features and benefit of ***SEWSS*** along with an example of its database integration capabilities.

For more details about ***SEWSS*** capabilities and for a customized demonstration of the system for your needs, please contact StatSoft at 918-749-1119.

This paper, "Part 1: Database Integration", is the first in a series that address the database integration capabilities of ***SEWSS***. Please refer to the other two companion papers, entitled "Part 2: Interactive Filtering" and "Part 3: Applying Specifications".