Viscovery® SOMine 7 – What's New?

New in Viscovery® SOMine 7 Data Mining Suite

New architecture

The new **Viscovery® SOMine 7** data mining suite contains the complete feature set of the previous standalone tools of the Viscovery 6 product family. All functions formerly provided by these products – Viscovery Profiler, Viscovery Predictor, Viscovery Scheduler, and Viscovery SOMine 6 – are now accessible through workflows of **Viscovery® SOMine 7**. The suite consists of a basic module, **Visual Clusters**, which can be operated as a standalone tool, and several extension modules, which allow the system to be flexibly expanded as needed.

- The minimum configuration of **Viscovery® SOMine 7**, **Visual Clusters**, provides streamlined navigation through the fundamental data mining functions based on self-organizing maps.
- For integrated exploration, predictive modeling and model application, the **Explore and Classify** and **Predict and Score** extension modules are available.
- The **Enterprise Data** extension module provides import, handling, and analysis of high-dimensional and voluminous data and the ability to export and import data preprocessing settings from/to spreadsheets.

Viscovery® SOMine 7 projects now accommodate up to five workflows, depending on the licensed modules. Each workflow can be automated using the corresponding **Workflow Automation Service**, which is available as part of an additional package. As before, predictive models can be integrated in operational systems for real-time application using the **Viscovery® One(2)One Engine**.

Streamlined workflows

The former Viscovery 6 workflows have been unified and renamed to accommodate exploration, classification, and prediction functions in a single project structure. In particular, the former workflows are now available as the following:

- **Create Data Mart** has been renamed **Preprocess**
- **Create Model** of SOMine and Profiler has been renamed **Cluster**
- **Apply Model** of SOMine and Profiler has been renamed **Classify**
- **Evaluate** of SOMine and Profiler has been functionally integrated in the new **Classify** workflow
- **Create Model** of Predictor has been renamed **Predict**
- **Apply Model** of Predictor has been renamed **Score**
- **Evaluate** of Predictor has been functionally integrated in the **Score** workflow

The **Preprocess** and **Cluster** workflows are now part of the basic **Visual Clusters** module. The **Classify** workflow comes with the **Explore and Classify** extension module; the **Predict** and **Score** workflows are included in the **Predict and Score** module. More detailed descriptions of the new workflow features are provided in the sections for the corresponding modules below.

There is now only one type of project, which can contain all types of workflows – **Preprocess, Cluster, Classify, Predict,** and **Score**. These projects are now stored in *.visdm files. Earlier Viscovery project file formats (*.vsp, *.csp, and *.vpp from version 4, 5, and 6) of Viscovery SOMine, Viscovery Profiler, and Viscovery Predictor can be imported into Viscovery 7 projects.

Viscovery® SOMine 7 – What's New?

Modular licensing

For each modular configuration of **Viscovery® SOMine 7**, the following license options regarding the term and the user base are available:

- Annual term-licenses, which include minor updates for one year, or perpetual licenses
- Single-user licenses or network licenses for multiple users within local networks

Each licensed user may operate one interactive instance of the software at a time. Single-user licenses are bound to the user account on a single computer and may be transferred to a different computer system once per year. Network licenses allow operation of the software for a defined maximum of concurrent users and require the additional installation of the new **Viscovery® License Server**.

Upgrading previous Viscovery software versions to version 7

The correspondence between earlier versions of Viscovery software and the new modules in Viscovery 7 is summarized in the following table.

Viscovery version 6 and earlier	**Viscovery SOMine 7 module combination**
Viscovery Profiler	**Visual Clusters**, **Explore and Classify**, and **Enterprise Data**
Viscovery Predictor	**Visual Clusters**, **Predict and Score**, and **Enterprise Data**
Viscovery Scheduler	**Workflow Automation Services***
Viscovery SOMine Basic Edition	**Visual Clusters****
Viscovery SOMine Expert Edition	**Visual Clusters** and **Explore and Classify**
Viscovery SOMine Enterprise Edition	**Visual Clusters**, **Explore and Classify**, and **Enterprise Data*****

* Viscovery Scheduler allowed automation of all Viscovery Profiler and Viscovery Predictor workflows, whereas each Workflow Automation Service is bound to a single workflow of Viscovery SOMine 7.

** This upgrade path does not include the following functions available in the previous Basic Edition: creation of alternative segmentations directly in the model, application of a model to new data and evaluation of applications.

*** This upgrade path additionally enables database access and optional model automation and integration, making it functionally equivalent to the former Viscovery Profiler.

Viscovery® SOMine 7 – What's New?

New in Visual Clusters Basic Module

Data preprocessing

- The facility to define new attributes in the **Preprocess** workflow has been moved from the first workflow step to the second workflow step (which has been renamed to "Define Attributes"). Thus, data do not need to be reimported when attribute definitions are modified in alternative workflow branches.
- The order in which the imported data attributes are stored can be defined earlier in the **Preprocess** workflow (in the "Define Attributes" step), not only at the "Write Data Mart" step, which is the last step.

Cluster analysis and SOM exploration

- A preferred cluster method and an initial number of clusters can be selected prior to map creation so that the corresponding clustering is available in the computed SOM. Alternative cluster models can be defined in the New Segmentation dialog available from the **Explore and Classify** module.
- Access to external documents from SOM pictures and thumbnail popups over SOM pictures no longer requires that the Data Records window is visible.

Data export

- Data exported to text files are now encoded in UTF-8 to allow for the use of national characters.
- Data can now be exported also to *.xlsx Microsoft Excel files.

New in Explore and Classify Extension Module

SOM model exploration

- The Group Profile window now displays the p-value of the statistical test used to compare groups instead of the confidence level (1−p).
- The ability to specify "action codes" for clusters has been removed to streamline the user interface.
- Records from a time-ordered data file can be displayed as animated process trajectories in the SOM model.

Application of classification models

- When new data is classified in the **Classify** workflow, the control group percentage is now specified only once and is applied to all clusters; the percentage is no longer specified separately for each cluster.
- The ability to record a sequence of comments with segmentations was removed in favor of a streamlined overview of segmentations.
- The meaning of the control group value that is imported and exported in the **Classify** workflow has been inverted: the value 1 now identifies records that belong to the control group and the value 0 identifies records that belong to the application group.

- The application name does not need to be specified when the classification results are exported to a file. When exported to a database, an application name is still required.

Evaluation of classification results

- Means and sums of attributes for each cluster can now be inspected in a chart accessible from the **Classify** workflow. The chart offers better interactive attribute selection than was possible in the former **Evaluate** workflow.
- The chart formerly used to assess the quality of the classification (previously available in the **Evaluate** workflow of Viscovery Profiler and Viscovery SOMine) is now integrated in the chart of the Check Application step of the **Classify** workflow.
- The model's classification performance can now be compared to the performance of a random classifier when the real class names of data records are known.

New in Predict and Score Extension Module

Model creation and definition

- Partitioning of a model data mart into model data and test data is executed in the **Predict** workflow rather than at the beginning of the **Preprocess** workflow, as it was in Viscovery 6.
- The local regression algorithm now produces fewer outliers in local regression coefficients when a variable has only a few distinct values, as is common for nominal variables.
- If the **Explore and Classify** module is included in the license, the corresponding features are also available to cluster and profile the resulting map of the local regression model.

Application of scoring models

- When a prediction is applied to new data (**Score** workflow), the original attribute values can optionally be copied to the result file or database table.
- The meaning of the control group value that is imported and exported in the **Score** workflow is now inverted: the value 1 now identifies records that belong to the control group and the value 0 identifies records that belong to the application group.
- The scoring name does not need to be specified when the scoring results are exported to a file. When exported to a database, a scoring name is still required.

Model validation, application evaluation

- In the Validate Model chart, the model corresponding to the current workflow branch is indicated with a thick line.
- The two charts used to assess the quality of the scoring application in the previous **Evaluate** workflow of Viscovery Predictor are now consolidated into a single chart, which is integrated into the chart of the Check Application step of the **Score** workflow.
- The chart of the Check Application step now allows comparison of actual values and predicted values for the application, control group, or entire data.

Viscovery® SOMine 7 – What's New?

New in Enterprise Data Extension Module

Data import and export

- When a classification or scoring result is exported to a database, arbitrary attribute values can be copied from the application data mart to the database table.
- SOM_PAK model files (*.cod) can now also be imported.

Join functionality

- Multiple data files and/or database tables can be joined directly in the "Import Data" step of the **Preprocess** workflow using a "left outer" join operation.
- Data marts need not include key attributes. However, if data sources are joined in the **Preprocess** workflow, key attributes must be present in the data tables to enable the join operation.

Preprocessing protocol

- Preprocessing settings of the **Preprocess** workflow can be imported from and exported to a spreadsheet document, which is called a "preprocessing protocol". This file provides a compact overview of all preprocessing parameters and can be edited outside the Viscovery system.
- The arrangement of preprocessing parameters in a table with columns of predefined content allows variable definitions to be specified, variable ranges to be set, the treatment of outliers and irregularities to be defined, and variable descriptions to be added.

New in Workflow Automation Services Package

Scheduling of tasks

- The user interface to schedule workflows is now integrated in Viscovery SOMine and is no longer a separate program.
- Scheduled workflow tasks can be cleared for immediate execution ("Start now"), so that the task is started before the scheduled time.

Automated execution of workflows

- The Workflow Automation Service runs in the background only if there are scheduled workflow tasks.
- Scheduled services run on behalf of the user, but do not require the user to be logged in on the computer.
- When workflow tasks must access network resources, the user's permissions are used. Therefore, it is necessary to enter the account password when tasks are scheduled.

New in Viscovery One(2)One Engine

New license options

- In accordance with Viscovery SOMine, user-base options for the One(2)One Engine now include single-instance licenses or network licenses for a defined maximum of concurrent instances.
- Network installations enable load balancing and fail-safe operation of the entire system.
- Network licenses allow installation of the software on an arbitrary number of computers and up to the defined maximum of instances to be run concurrently.
- Each opened model constitutes a running instance of the software until the model is closed.
- Network licenses require the additional installation of the Viscovery License Server.
- Communication timeouts between the Viscovery License Server and the Viscovery One(2)One Engine can be customized.