

## Welcome to the first KNIME Newsletter!

The KNIME community has grown quickly over the last few years. From a few hundred downloads of KNIME 1.0 in July 2006 there are now thousands of people around the world actively using KNIME. The KNIME Newsletter is our newest vehicle to support this growing community—every quarter we will report on a particularly interesting application of KNIME in the real world and complement this with a series of tips and tricks around the practical usage of KNIME.

We hope you enjoy reading this and future versions of KNIME News!

The KNIME Team

## Reports with KNIME: a success story

People working in the corporate world and, in particular, people from the financial world, well know just how many reports are required every year, every month, and even every day for the company to work properly. Communication across departments and between employees and management is crucial if the bank's activities are to run flawlessly. Some reports are required on demand while others are needed on a regular basis, depending on the specific nature of their content. An automatic scheduling solution is often required for regularly culled reports in

order to free up some of the company resources. Reports must also be repeatable. Business people in banks and other companies need reports that integrate data from diverse data sources. This frequently poses a technical problem. Data is often spread across one or more databases, a few Excel sheets of different structures here, some text files there, and maybe even employees' written notes.

Bank Frey is a medium-sized Swiss private bank. It was founded in 2000 with the strong commitment to secure and grow wealth for generations to come. Thanks to its talented team of top professionals with an entrepreneurial focus, Bank Frey has grown rapidly in the last 10 years. This rapid growth has made evident the need for a reliable and scalable report platform to cover all internal and external communication requirements.

Bank Frey was on the lookout for a Business Intelligence (BI) and reporting tool with low license costs and minimum training requirements that would work across a number of heterogeneous data sources and provide the scalability to handle growing amounts of data. After a rapid evaluation of the BI tools available on the market, Bank Frey decided to test the KNIME Reporting tool.



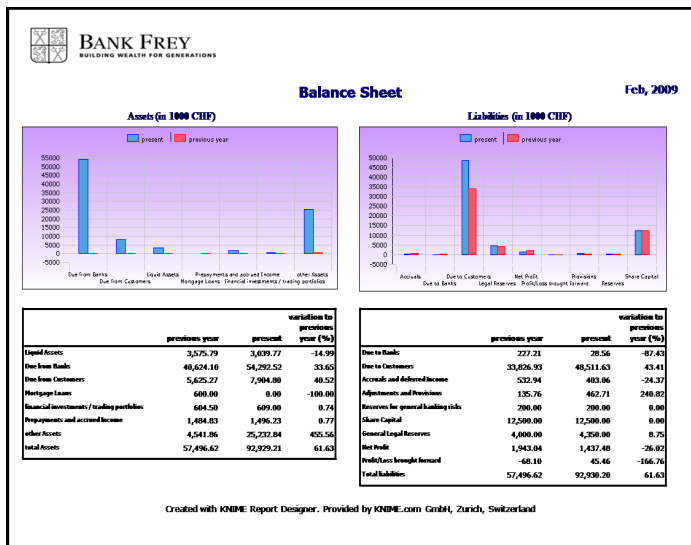
**Gregor Bienz**  
CEO  
Bank Frey  
Zurich  
Switzerland

Unlike many other BI tools, KNIME, being open source, can be fully tested free of charge. All it took was to download it, install it and run it. Its graphical interface is intuitive and does not require any previous knowledge of script or language programming. The full range of BI functionalities is extensive, spanning from the calculation of classical statistical variables to data manipulation functions to transform and re-arrange data: some basic knowledge of statistics is still needed, but, this provided, the transition to KNIME is quick and painless.

After only a few months of work, Bank Frey has set up a simple Data Warehousing (DWH) solution, involving a number of reports run on either a regular monthly or daily basis and further reports that are run on demand. The DWH workflows integrate data from Excel files, text files and databases and store them in a centralized location. The Report workflows read from this central location, perform the required data selection, and write the results into a report with a professional graphic layout. The DWH and the

**"Data from different sources is consolidated and is now available at any moment for reporting and analytics."**

on demand while others are needed on a regular basis, depending on the specific nature of their content. An automatic scheduling solution is often required for regularly culled reports in



Example of the reports running at Bank Frey.



## 4th KNIME User Meeting

### User Group Meeting (UGM)

Meet other KNIME Users and the entire KNIME Team!

**March 2—3, 2011**  
Technopark  
Zurich, Switzerland

[www.knime.org/ugm](http://www.knime.org/ugm)

If you also want to learn more about how to use KNIME and KNIME Reporting, you can enroll for the associated training courses.

### KNIME User Training

**Feb. 28—March 1, 2011**  
Technopark  
Zurich, Switzerland

## Next Developer Training

During this course you will be given a comprehensive overview of the KNIME architecture and learn how to extend KNIME by programming your own custom nodes.

**April 5-6, 2011**  
Technopark  
Zurich, Switzerland

Visit:  
[www.knime.org/training](http://www.knime.org/training)  
to register and for more information.

## KNIME 2.3 is out!

*Even Santa Claus can't mine without KNIME!*

Download the new KNIME 2.3 from:  
[www.knime.org/download](http://www.knime.org/download)  
and start mining your data today!



report workflows are integrated into the KNIME Server, meaning they are now available to more than one user and can be scheduled to run at convenient times. Access to and execution of DWH and reporting workflows are regulated by the user permissions feature on the KNIME Server. Reports on the server

can be executed from any web browser. To summarize, Bank Frey now has a small BI group working with KNIME. The cost of setting this up was not excessive: the training required by this BI group to become familiar with KNIME and its BI functions was minimal and the license costs for the

KNIME Server affordable. The DWH solution enables data from different sources to be consolidated and a history of all data coming into the bank is kept. These data are subsequently available at any time for analysis, regular adjustment or repetition of the generated reports.

## What's new in KNIME 2.3

On the evening of December 5, children in northern Europe put their shoes outside the front door. They know that overnight, Nikolaus (Santa Claus) will stop at every door to fill the shoes of children who have been good with all sorts of sweets for them to find in the morning.

Late last year Nikolaus has taken a wide detour and stopped at the doors of all KNIME users to fill their shoes with a new more powerful version of KNIME: KNIME 2.3. We have largely retained the familiar look and feel of KNIME in this release and added a few new interesting features that simplify and empower our daily data analysis work. This section of the newsletter is dedicated to the new features of version 2.3.

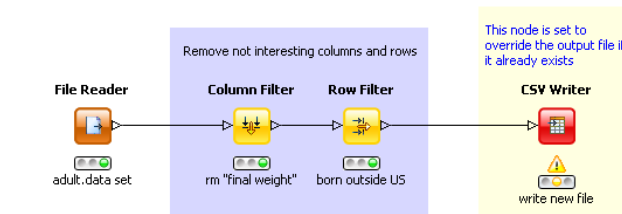
KNIME 2.3 can be downloaded from:

[www.knime.org/download](http://www.knime.org/download)

Shift+F7	executes all executable nodes
F7	executes selected and executable nodes
Shift+F9	cancels all running nodes
F9	cancels selected and running nodes
Ctrl+S	saves the workflow

### Hotkeys

For keyboard lovers and in general for all those users suffering from carpal tunnel syndrome, most KNIME commands can now run via hotkeys. All hotkeys are listed in the KNIME menus or in the tooltip messages at the side of their corresponding commands.



This workflow is my first KNIME workflow. It reads data, removes uninteresting columns and rows from the data table and writes the result to a CSV file.

### Workflow Annotations

It is now possible to include workflow annotations to comment and explain what each node or group of nodes does. Just right-click anywhere in the workflow editor and select "New Workflow Annotation". Then type in your explanatory comment, resize the annotation window to make clear which group of nodes it refers to, and there you have the explanation of this part of your workflow! The result is an improved overview of the whole workflow task.

## Tips and Tricks

### Label Data Sets

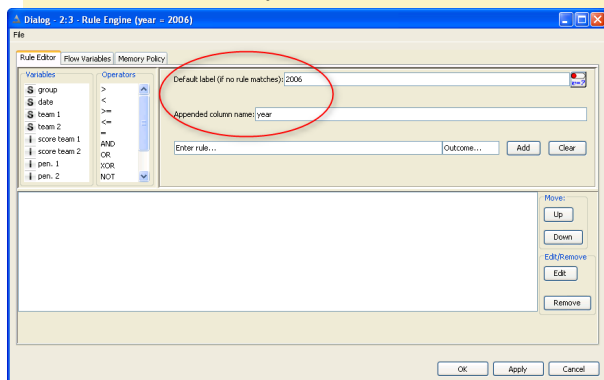
Very often we need to concatenate a number of data sets to feed the data analysis cycle. However, once the data analysis has been performed, we might want to recreate the original data subsets exactly as they were.

Recently, for example, I created a data set with the game scores from the 2006 Football World Cup and a data set with the game scores of the 2002 Football World Cup from the data available on the FIFA web site ([www.fifa.com](http://www.fifa.com)). In order to perform an analysis on both world cups' scores, I concatenated the two data sets into one big data set. I also wanted to display the scores of each world cup, 2002 and 2006, in two separate report tables.

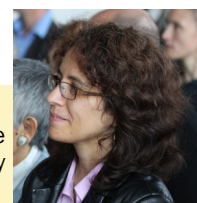
Before embarking on a cumbersome operation to identify the original data set of each data row by parsing the game date, I opted for a simpler solution: I assigned the world cup's year as an additional label to the data rows of each data set before concatenation.

To assign labels to data rows I often use the "default" label of the "Rule Engine" node. So, I used two "Rule Engine" nodes to create a new column "year" with default label "2002" in one data set and default label "2006" in the other data set. After

concatenation, the two original data sets can still be easily recovered based on the value of the column "year".



Definition of the default label "2006" in the new column "year" in the "Rule Engine" node.



**Rosaria Silipo**  
Data Mining Consultant  
bizzeps AG  
Zurich, Switzerland

### Identify duplicates

How many times have you created a unique key that turned out not to be unique at all? It happens quite often. The problem is: how do we verify that a unique key is actually unique? I use the "RowID" node.

The "RowID" node replaces the RowIDs of the input data with the values of another column, assuming that such values are unique. Indeed, if the checkbox "Ensure uniqueness" is NOT enabled, the "RowID" node will fail on non-unique values and the error message describes the first non-unique encounter as:

"ERROR RowID Execute failed: Encountered duplicate row ID "<value>" at row number <row\_number>"