



# KNIME Meetup Pavia/Milano

*Thomas Gabriel*

*KNIME.com AG, Zurich, Switzerland*

# A Brief History of KNIME

- 2004: KNIME development commences
- 2006: KNIME v1 released
- 2006: Spin-off in Konstanz, Germany
- 2006-2007: First commercial partners
- 2008: KNIME moves to Zurich
- 2010: Enterprise products released
- 2011: KNIME.com AG founded
- 2013: KNIME opens San Francisco office...



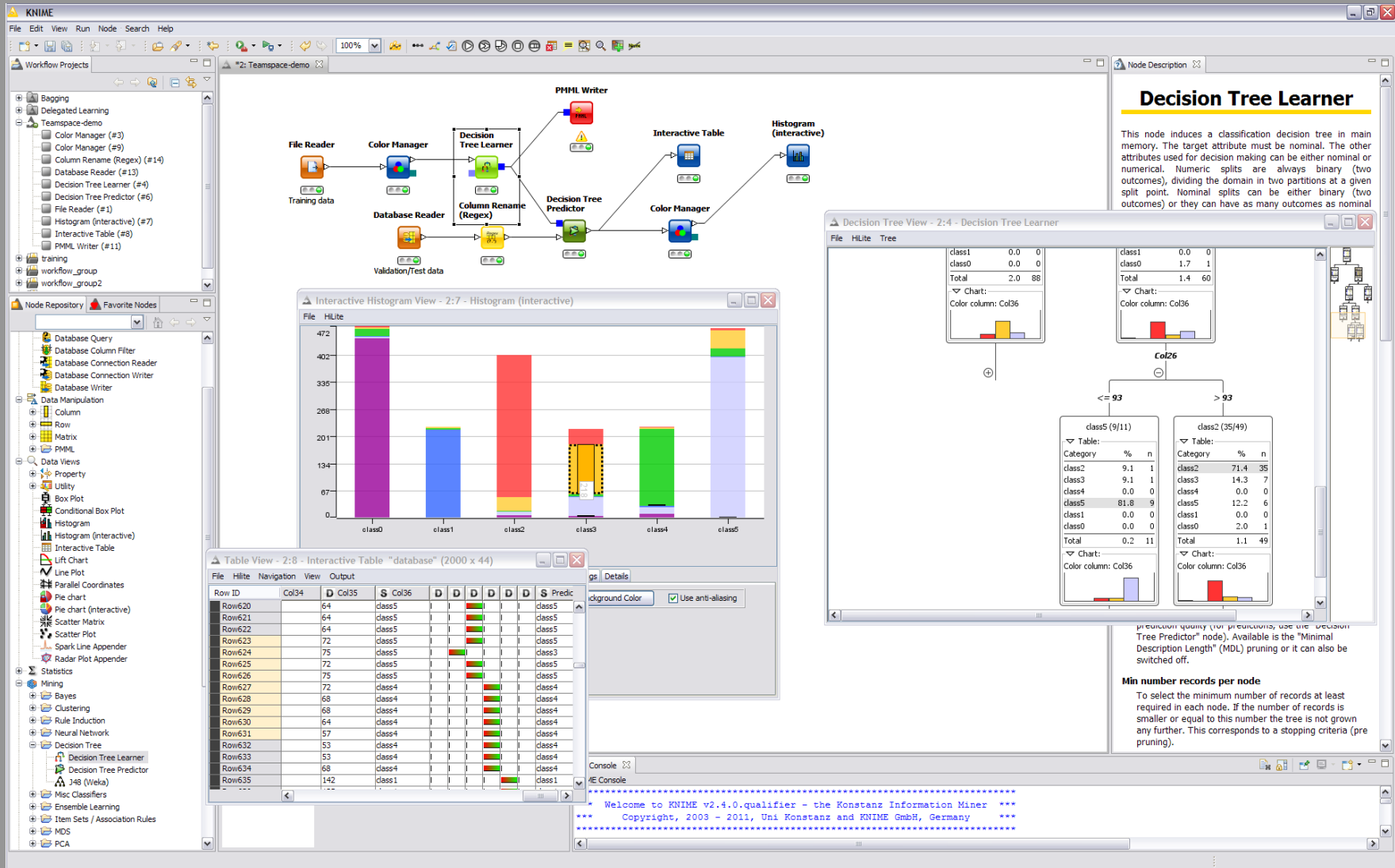
## Status Quo:

- +3000 Organizations Using KNIME
  - ~30% Life Science
  - ~70% Business Intelligence, Analytics
- +50 Very Active Community Developers

## KNIME.com AG

Turnover:	Small but profitable!
Staff:	20+

# The KNIME Platform



The screenshot displays the KNIME platform interface with a workflow titled "Teanspace-demo". The workflow includes nodes for File Reader, Color Manager, Database Reader, Decision Tree Learner, Column Rename (Regex), Decision Tree Predictor, PMML Writer, Interactive Table, and Histogram (interactive).

**Decision Tree Learner Node Description:**

**Decision Tree Learner**

This node induces a classification decision tree in main memory. The target attribute must be nominal. The other attributes used for decision making can be either nominal or numerical. Numeric splits are always binary (two outcomes), dividing the domain in two partitions at a given split point. Nominal splits can be either binary (two outcomes) or they can have as many outcomes as nominal

**Interactive Histogram View - 2:7 - Histogram (interactive)**

File HILite

class	0.0	0
class1	0.0	0
class0	0.0	0
Total	2.0	88

**Table View - 2:8 - Interactive Table "database" (2000 x 44)**

Row ID	Col34	D Col35	S Col36	D	D	D	D	D	D	S	Predic
Row620	64	class5									class5
Row621	64	class5									class5
Row622	64	class5									class5
Row623	72	class5									class5
Row624	75	class5									class3
Row625	72	class5									class5
Row626	75	class5									class5
Row627	72	class4									class4
Row628	68	class4									class4
Row629	68	class4									class4
Row630	64	class4									class4
Row631	57	class4									class4
Row632	53	class4									class4
Row633	53	class4									class4
Row634	68	class4									class4
Row635	142	class1									class1

**Decision Tree View - 2:4 - Decision Tree Learner**

File HILite Tree

class	0.0	0
class1	0.0	0
class0	1.7	1
Total	1.4	60

Chart: Color column: Col36

Col26

<= 93      > 93

class5 (9/11)

Category	%	n
class2	9.1	1
class3	9.1	1
class4	0.0	0
class5	81.8	9
class1	0.0	0
class0	0.0	0
Total	0.2	11

Chart: Color column: Col36

class2 (35/49)

Category	%	n
class2	71.4	35
class3	14.3	7
class4	0.0	0
class5	12.2	6
class1	0.0	0
class0	2.0	1
Total	1.1	49

Chart: Color column: Col36

prediction quality (for predictions, use the "Decision Tree Predictor" node). Available is the "Minimal Description Length" (MDL) pruning or it can also be switched off.

**Min number records per node**

To select the minimum number of records at least required in each node. If the number of records is smaller or equal to this number the tree is not grown any further. This corresponds to a stopping criteria (pre pruning).

Console

```

*** Welcome to KNIME v2.4.0.qualifier - the Konstanz Information Miner ***
*** Copyright, 2003 - 2011, Uni Konstanz and KNIME GmbH, Germany ***
  
```

KNIME loads and integrates data from diverse data sources:

- Different data bases
- Various file formats (CSV, XML, SDF, etc.)

#### File Reader



Excel import

#### Database Connector



Node 0: 1:8

#### PMML Reader



Vendor independent predictive model

KNIME provides huge repository of modules for easy-to-use, modular

- Data preprocessing
- Data fusion
- Data transformation

File Reader



Excel import

Missing Value



Node 0:1:20

Pivoting



Node 0:1:7

Database Connector



Node 0:1:8

Database Column Filter

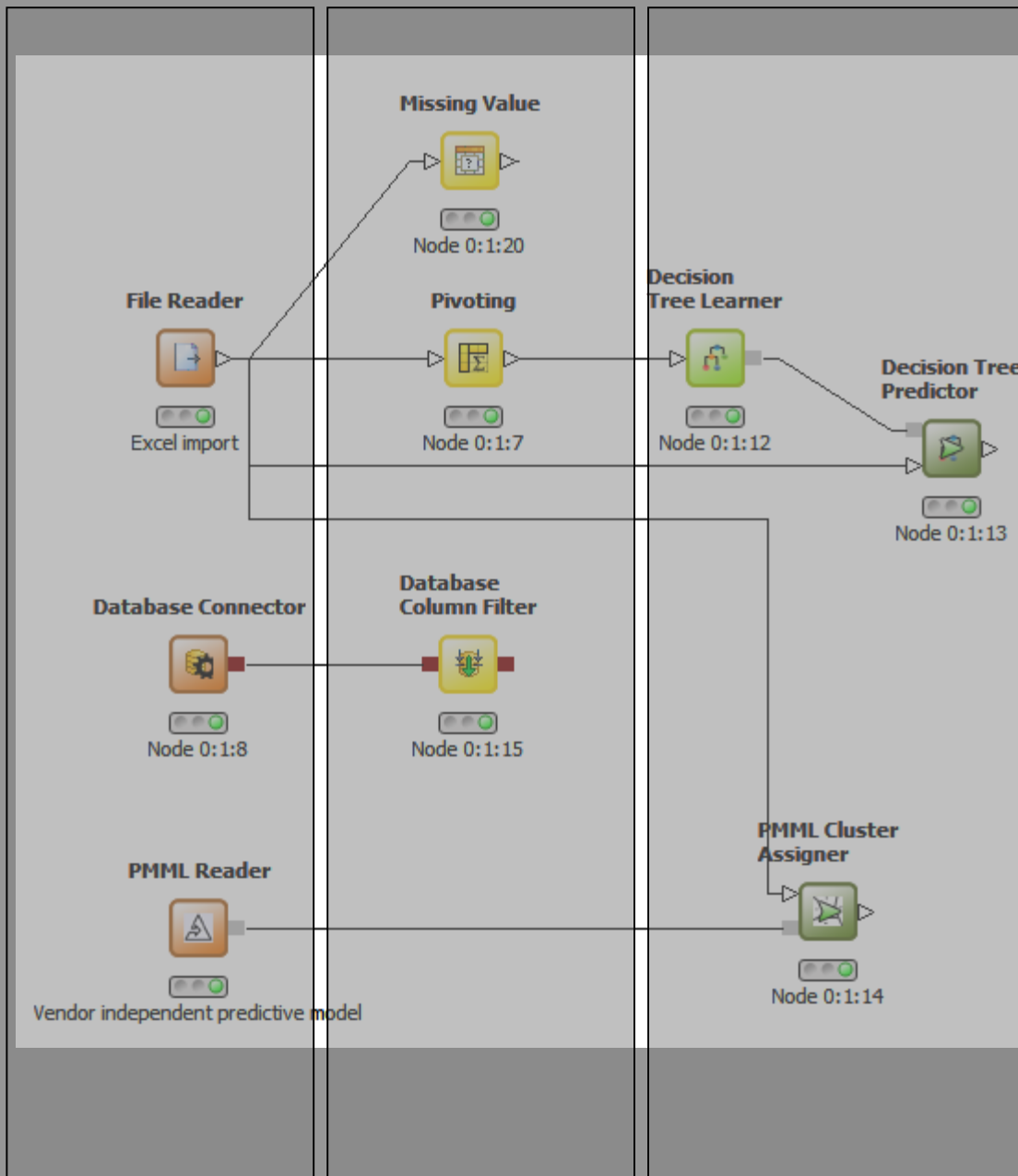


Node 0:1:15

PMML Reader



Vendor independent predictive model



In addition to standard data mining techniques, KNIME adds cutting edge data analysis algorithms. (...thanks to its academic roots)

Missing Value



Node 0:1:20

Pivoting

File Reader

Decision  
Tree Learner

Interactive views provide data overviews and insights into the learned models.

Interactive linking & brushing techniques allow for powerful exploration of models and data.

Histogram



Node 0:1:18

Pie chart



Node 0:1:19

Node 0:1:8

PMML Reader



Vendor independent predictive model

Node 0:1:15

PMML Cluster  
Assigner



Node 0:1:14

Box Plot



Node 0:1:16

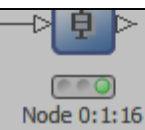
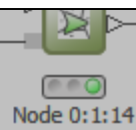
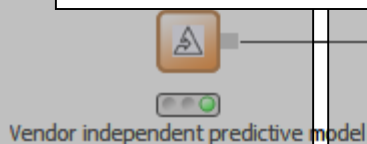
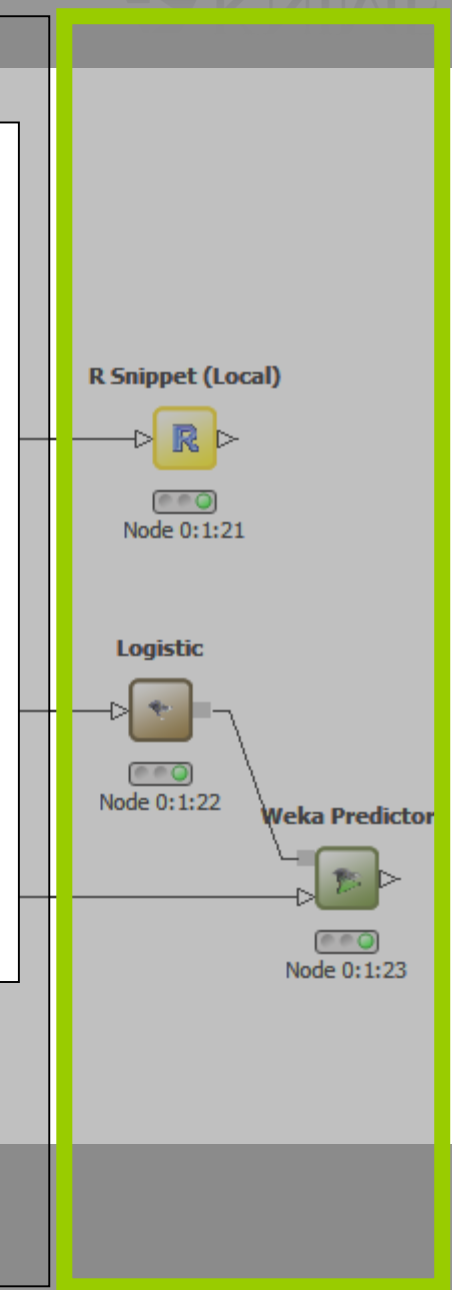
Due to its open API and “node-in-a-sandbox”-approach additional (also external) tools are easily integrated,

e.g.

- Access to the R Project (statistical analysis/visualizations)
- Complete integration of the machine learning library WEKA
- Application area specific integration, e.g. CDK (Chemical Development Kit), RDKit, ImageJ, ...

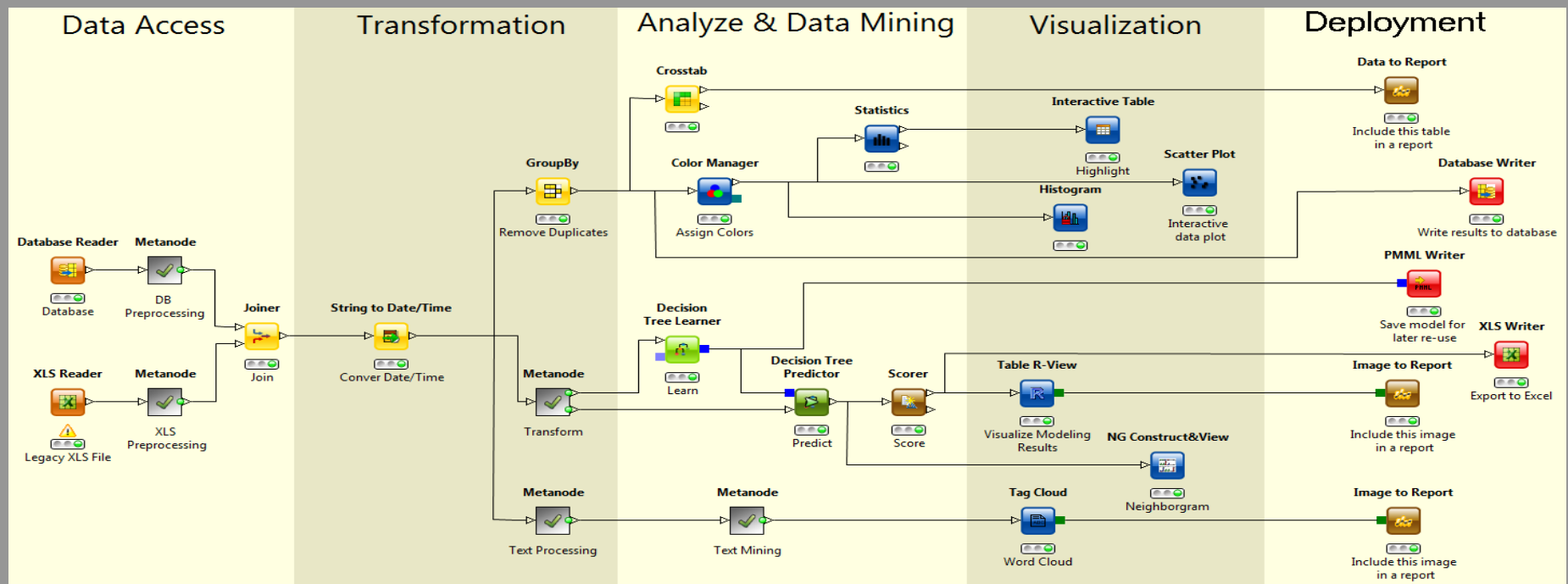


KNIME is Eclipse-based: Integrating other Eclipse projects such as BIRT, DTP, etc. provides even more functionality





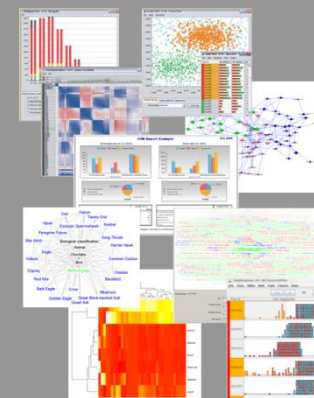
# Over 1000 native and embedded nodes included:



MySQL, Oracle, etc.  
 SAS, SPSS, etc.  
 Excel, Flat, etc.  
 Hive etc.  
 XML, PMML  
 Text, Doc, Image  
 Web Crawlers  
 Industry Specific  
 Community / 3rd

ETL  
 Row,  
 Column  
 Matrix  
 Text, Image  
 Time Series  
 Java  
 Python  
 Community / 3rd

Statistics  
 Data Mining  
 Machine Learning  
 Web Analytics  
 Text Mining  
 Network Analysis  
 Social Media Analysis  
 WEKA  
 R  
 Community / 3rd



R  
 JFreeChart  
 Community / 3rd

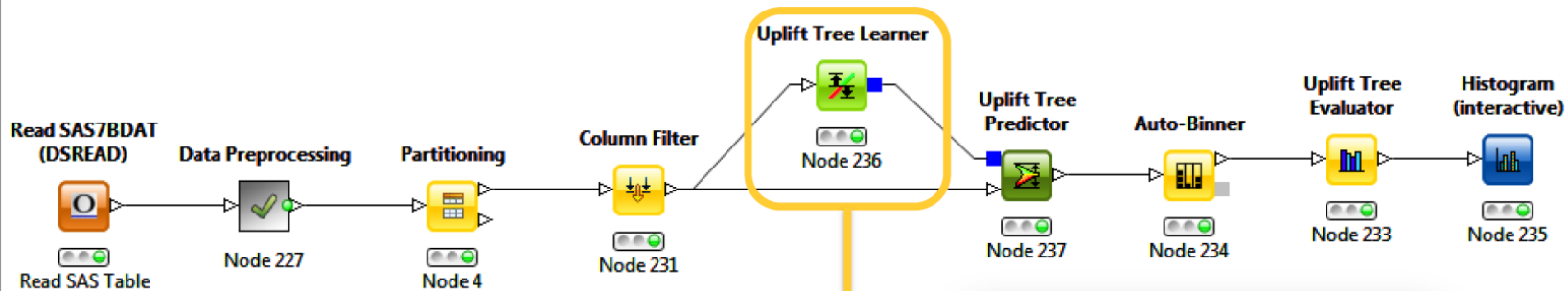
via BIRT  
 PMML  
 XML  
 Databases  
 Excel, Flat, etc.  
 Hive etc.  
 Text, Doc, Image  
 Industry Specific  
 Community / 3rd

# KNIME in Action: Standardization

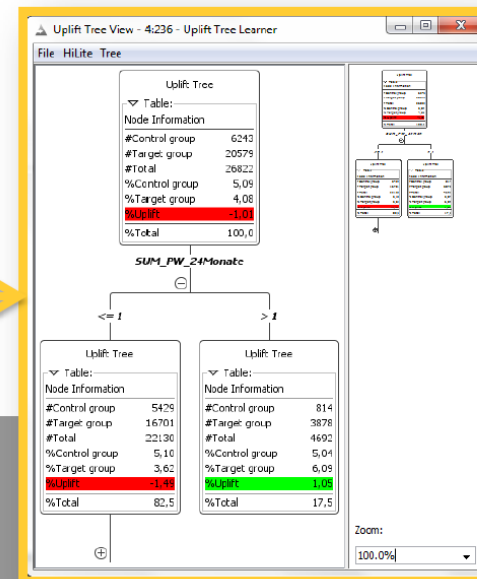
“[KNIME] helps me structure my procedures in a clean, reproducible, and auditable way that monster spreadsheets just can’t match...”



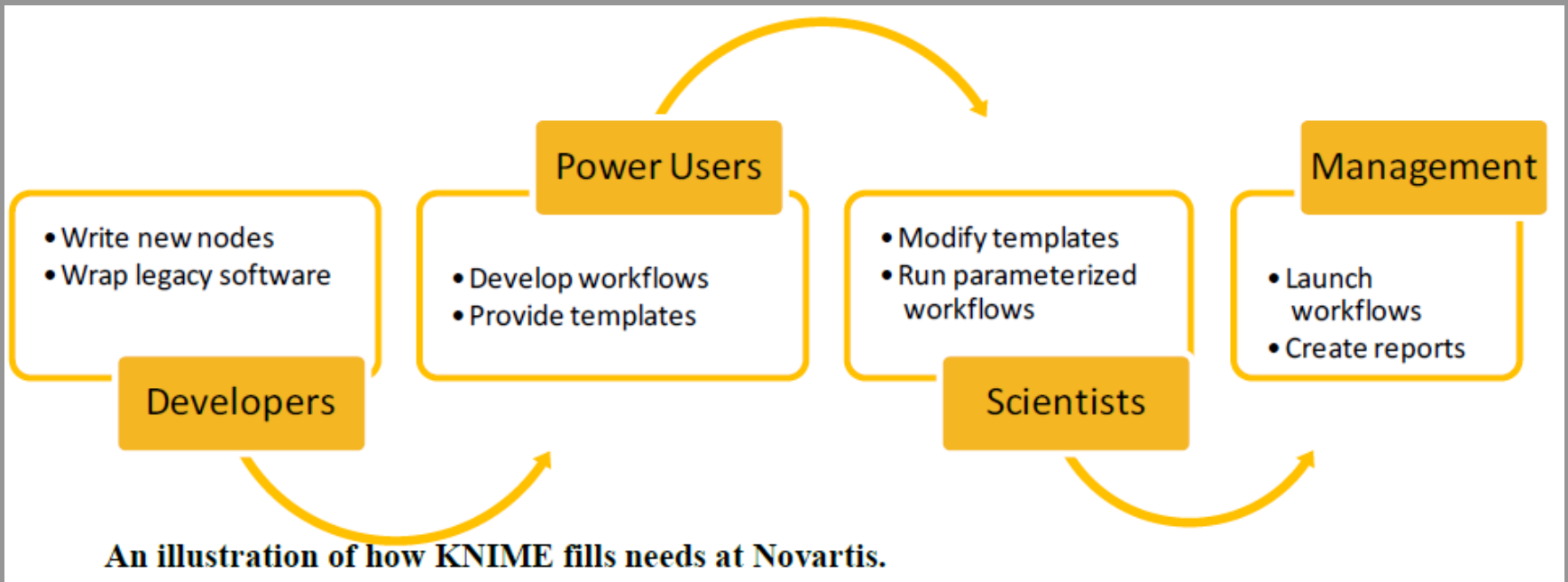
# Tool Integration



DYMATRIX Uplift Tree maximizes incremental impact by identifying the undecided clients that can be motivated by marketing.



# Life Science: Case Study





Mark Your Calendars:

KNIME's 7<sup>th</sup> User Group Meeting

12.-13. February 2014

Zurich, Switzerland