



# Integrating Data, Tools, and Science

*Michael R. Berthold*

*KNIME.com GmbH, Zurich, Switzerland*

*and*

*University of Konstanz, Germany*

# Agenda

KNIME: 2004 - today

or

How we went out to build a data analysis workbench and ended up building something else...

# Why KNIME?

# Why?

- University Group - Research Focus:
  - Data Mining / Machine Learning
  - Explorative Methods
  - „...Sparking Ideas“ (...PKDD/ECML 2005)
- Need for platform to develop, test, deploy, and use new methods.

# Requirements

- Modular & flexible
- Professional strength (industry collaborators)
- Highly scalable (dito)
- Easy to extend (legacy and third party tools)

**„I am working on KNIME“**

**(PhD Student, 2004/2005)**

„I am working with KNIME“

(PhD Students, 2006...)

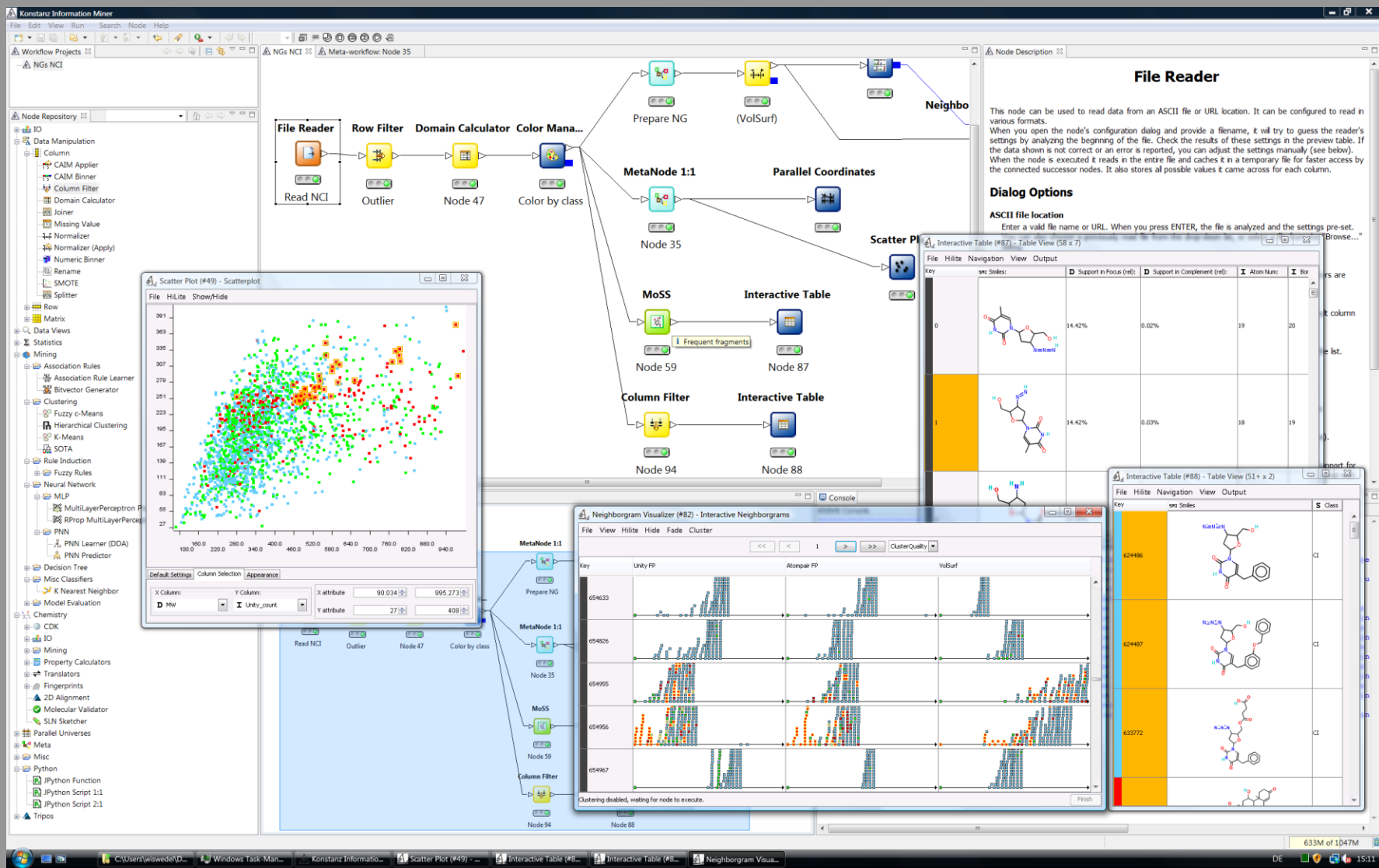
**June 2006:**

**First Open Source Release (v1.0)**



# What is KNIME?

# The KNIME Platform



The screenshot displays the KNIME platform interface with a workflow and several data visualization windows.

**Workflow:** The workflow starts with a **File Reader** node (Node 35) reading NCI data. This is followed by a **Row Filter** (Node 47) and a **Domain Calculator** (Node 47). The data then goes to a **Color Manager** (Node 35) which colors data by class. This is followed by a **MetaNode 1:1** (Node 35) which branches into several paths: **Prepare NG** (Node 59), **Parallel Coordinates** (Node 87), **MoSS** (Node 88), and **Column Filter** (Node 94). The **Column Filter** node is followed by another **Interactive Table** (Node 88).

**Scatter Plot (#49) - Scatterplot:** A scatter plot showing data points colored by class. The X-axis ranges from 100.0 to 940.0, and the Y-axis ranges from 27 to 391.

**File Reader Node Description:**

**File Reader**

This node can be used to read data from an ASCII file or URL location. It can be configured to read in various formats. When you open the node's configuration dialog and provide a filename, it will try to guess the reader's settings by analyzing the beginning of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below). When the node is executed it reads in the entire file and caches it in a temporary file for faster access by the connected successor nodes. It also stores all possible values it came across for each column.

**Dialog Options**

**ASCII file location**

Enter a valid file name or URL. When you press ENTER, the file is analyzed and the settings pre-set.

**Interactive Table (#87) - Table View (58 x 7):**

Key	Support in Focus (rel.)	Support in Complement (rel.)	Atom Num.	Bo
0	14.42%	0.02%	19	20
1	14.42%	0.03%	18	19

**Interactive Table (#88) - Table View (51 x 2):**

Key	Series	Class
624486	<chem>N#CC1=CC=C(C=C1)O</chem>	CI
624487	<chem>N#CC1=CC=C(C=C1)O</chem>	CI
633772	<chem>N#CC1=CC=C(C=C1)O</chem>	CI

**Neighborhood Visualizer (#82) - Interactive Neighborhoods:** A visualization showing neighborhood relationships for keys 654633, 654826, 654955, 654956, and 654967. It displays three columns: **Unary PP**, **Atompair PP**, and **VoSurf**.

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

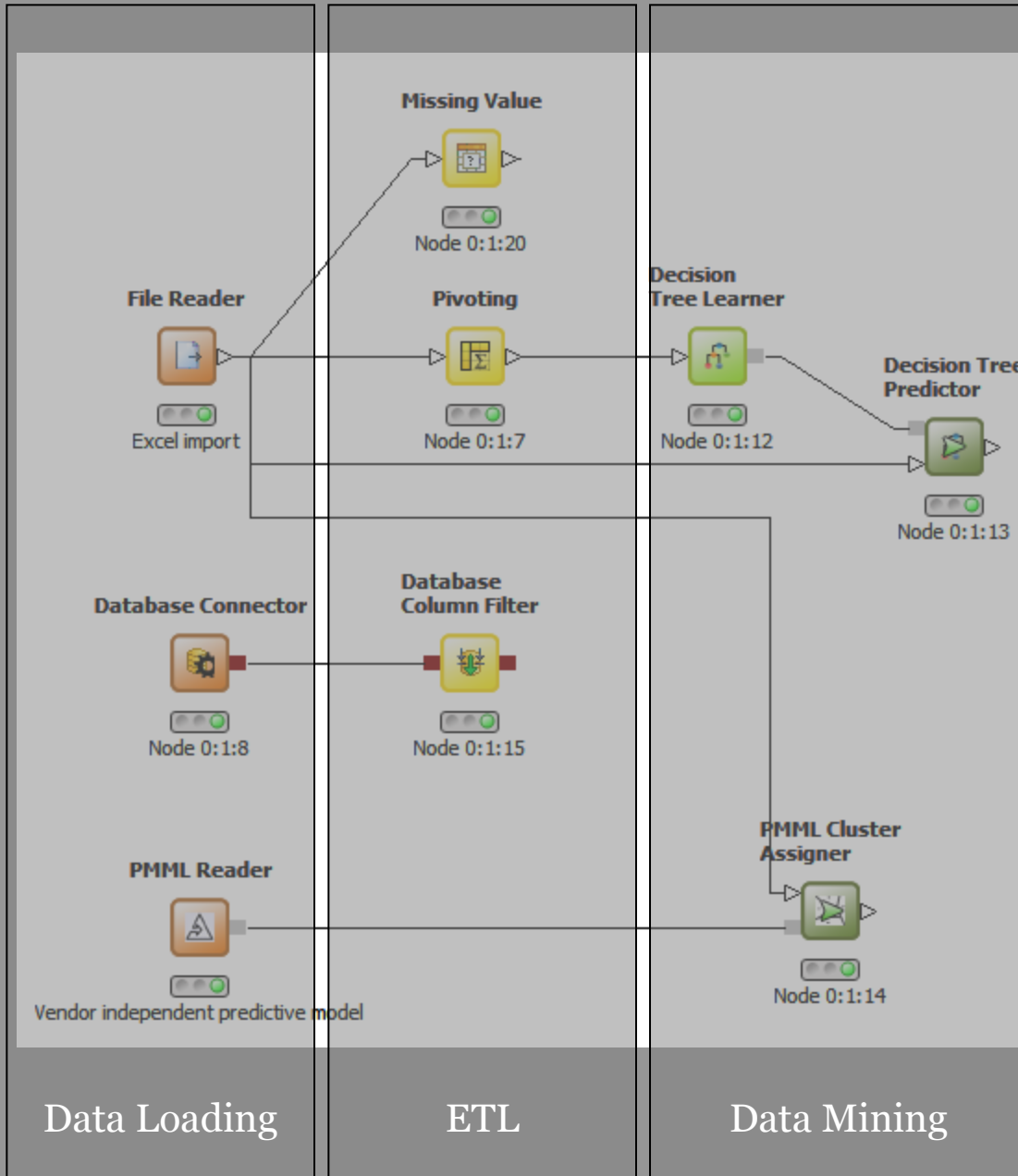
PMML Reader



Vendor independent predictive model

Data Loading

ETL



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

Data Loading

ETL

Data Mining

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

Data Loading

ETL

Data Mining

Visualization

# A peak under the hood

- Java
  - Runs on Linux, Windows, Mac
- Eclipse plugin
  - Extension wizard for own nodes
  - Simple installation of additional modules (text, image, time series, ... analysis)
- Out of memory data handling
  - Caches data to HD
  - Scales linearly (e.g. 200 - 20.000.000 images)
- Support for multi threaded execution
- Architecture designed by professional SW engineer

„But I have all those R scripts“

„Where is this algorithm?  
(Weka has it!)“

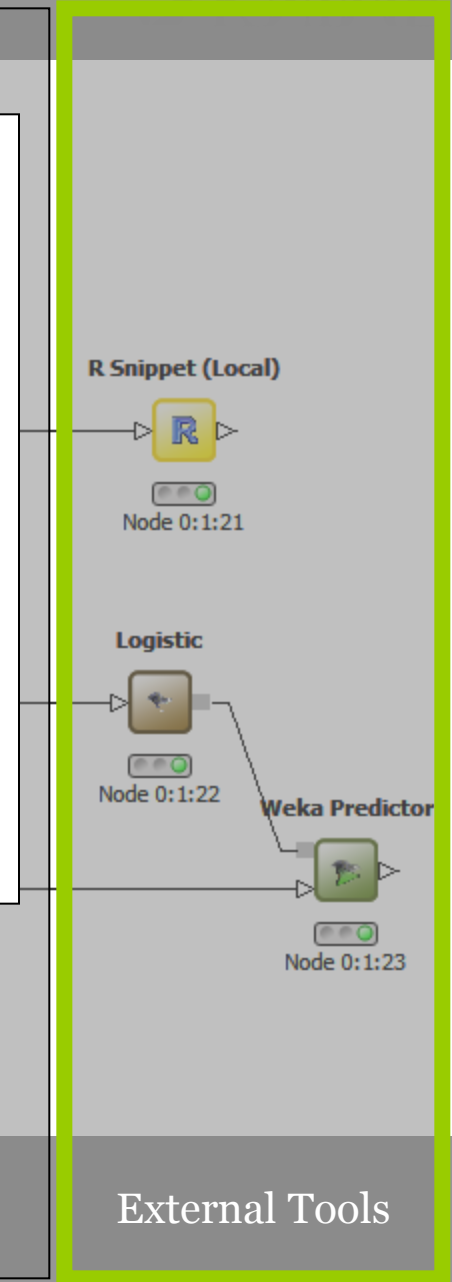
...re-invent the wheel?



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

- Access to the statistics tool R
- Complete integration of the machine learning library WEKA
- Application area specific integration, e.g. CDK (Chemical Development Kit)

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



PMML Reader  
Vendor independent predictive model

Assigner  
Node 0:1:14

Box Plot  
Node 0:1:16

R Snippet (Local)  
Node 0:1:21

Logistic  
Node 0:1:22

Weka Predictor  
Node 0:1:23

Data Loading

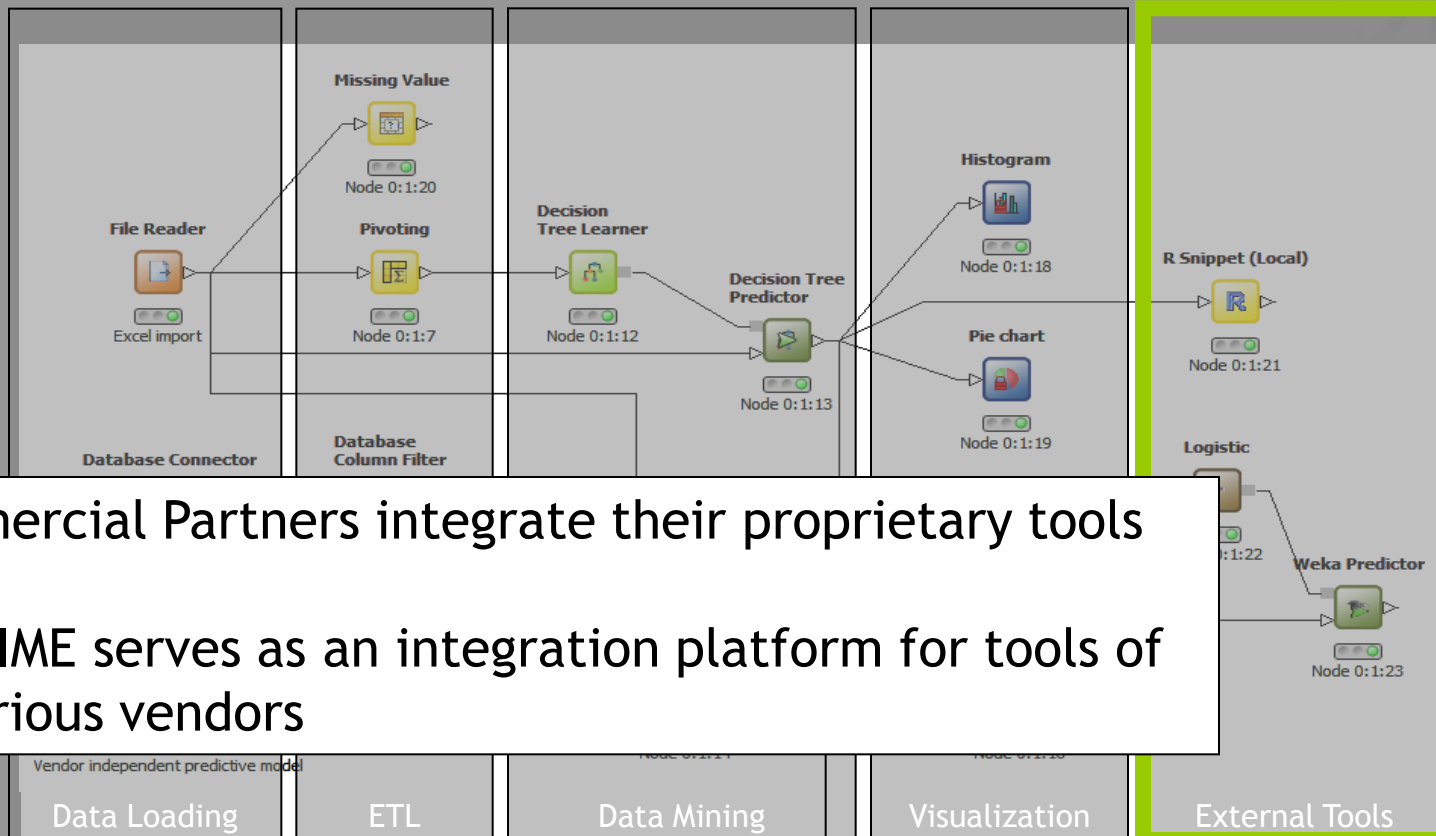
ETL

Data Mining

Visualization

External Tools

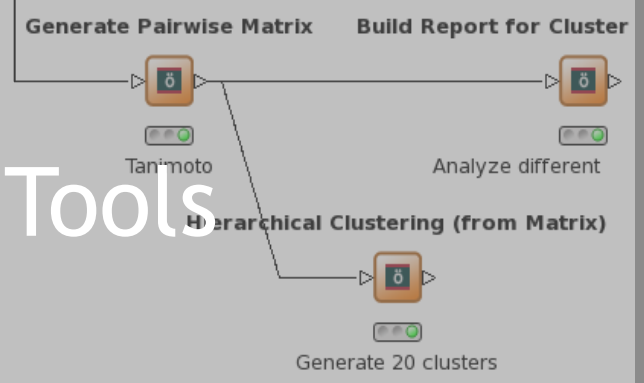
**„We bought this wonderful tool“**



Commercial Partners integrate their proprietary tools  
 ⇒ KNIME serves as an integration platform for tools of various vendors

Vendor independent predictive model | Data Loading | ETL | Data Mining | Visualization | External Tools

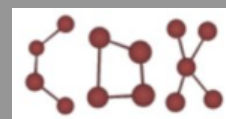
# 3rd Party Tools



# Some KNIME Partners



- CambridgeSoft
- Schrödinger
- Symyx Technologies
- Tripos
- ChemAxon (via Infocom)
- CDK
- Molecular Discovery
- Trewaren
- Tibco / Spotfire
- ...
- And others: BioSolve IT, CCG, ...



# Partner Extensions at Work

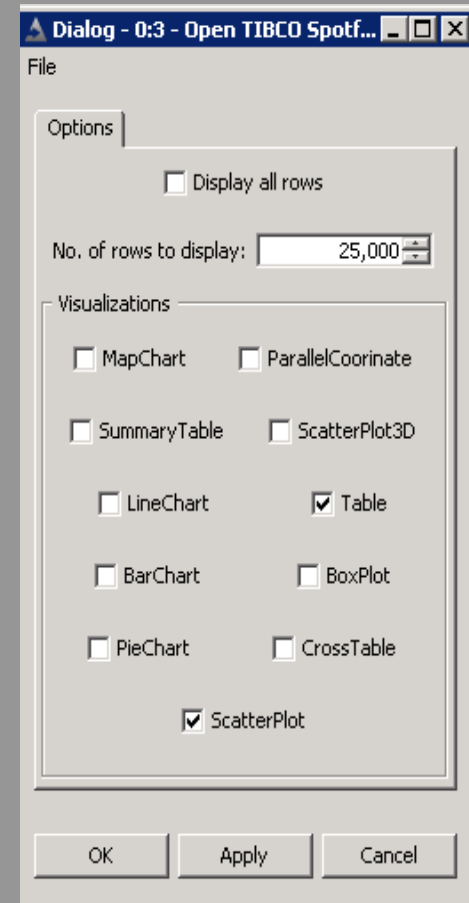
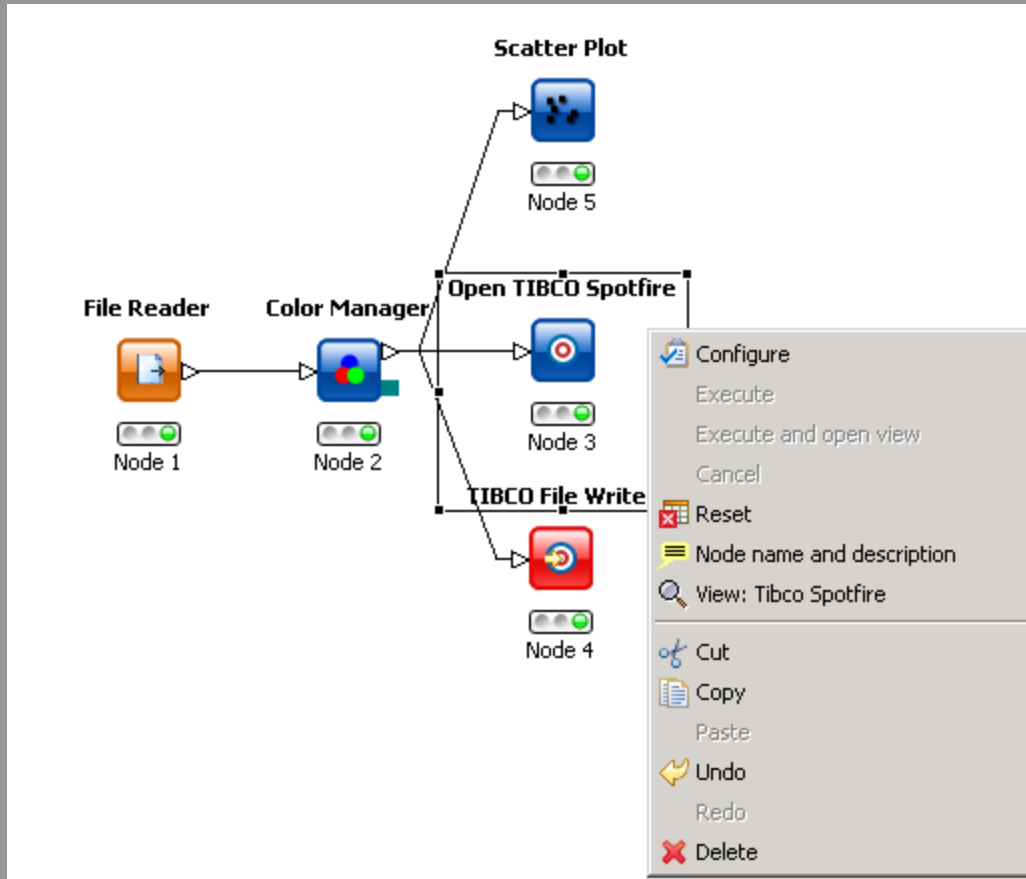
The screenshot displays the KNIME software interface with a workflow titled "Multi\_parameter\_vendor\_1". The workflow consists of several nodes: SDF Reader (Node 61), Molecular Descriptors (Node 54), 2D Properties (Node 45), MrvFromSdfConverter (Node 44), Polar Surface Area (Node 43), DBTranslate (Local) (Node 28), UNITY Fingerprint (Node 52), Keys (Node 39), Join Mole Properties (Node 58), and Prepare Descriptors. Orange arrows point from these nodes to five yellow boxes on the right, each representing a partner extension: Schrödinger, Tripos, Infocom, Symyx, and Weka, R, ...

Below the workflow, two data views are shown:

**Table View - 2:60 - Interactive Table (29 x 16)**

Row ID	sdf Molecule_input	EXTRE...	AlogP	MR
Row22	<chem>C1=CC=C(C=C1)N2C=CC(=C2)N3C=CC(=C3)N4C=CC(=C4)N5C=CC(=C5)N6C=CC(=C6)N7C=CC(=C7)N8C=CC(=C8)N9C=CC(=C9)N10C=CC(=C10)N11C=CC(=C11)N12C=CC(=C12)N13C=CC(=C13)N14C=CC(=C14)N15C=CC(=C15)N16C=CC(=C16)N17C=CC(=C17)N18C=CC(=C18)N19C=CC(=C19)N20C=CC(=C20)N21C=CC(=C21)N22C=CC(=C22)N23C=CC(=C23)N24C=CC(=C24)N25C=CC(=C25)N26C=CC(=C26)N27C=CC(=C27)N28C=CC(=C28)N29C=CC(=C29)N30C=CC(=C30)N31C=CC(=C31)N32C=CC(=C32)N33C=CC(=C33)N34C=CC(=C34)N35C=CC(=C35)N36C=CC(=C36)N37C=CC(=C37)N38C=CC(=C38)N39C=CC(=C39)N40C=CC(=C40)N41C=CC(=C41)N42C=CC(=C42)N43C=CC(=C43)N44C=CC(=C44)N45C=CC(=C45)N46C=CC(=C46)N47C=CC(=C47)N48C=CC(=C48)N49C=CC(=C49)N50C=CC(=C50)N51C=CC(=C51)N52C=CC(=C52)N53C=CC(=C53)N54C=CC(=C54)N55C=CC(=C55)N56C=CC(=C56)N57C=CC(=C57)N58C=CC(=C58)N59C=CC(=C59)N60C=CC(=C60)N61C=CC(=C61)N62C=CC(=C62)N63C=CC(=C63)N64C=CC(=C64)N65C=CC(=C65)N66C=CC(=C66)N67C=CC(=C67)N68C=CC(=C68)N69C=CC(=C69)N70C=CC(=C70)N71C=CC(=C71)N72C=CC(=C72)N73C=CC(=C73)N74C=CC(=C74)N75C=CC(=C75)N76C=CC(=C76)N77C=CC(=C77)N78C=CC(=C78)N79C=CC(=C79)N80C=CC(=C80)N81C=CC(=C81)N82C=CC(=C82)N83C=CC(=C83)N84C=CC(=C84)N85C=CC(=C85)N86C=CC(=C86)N87C=CC(=C87)N88C=CC(=C88)N89C=CC(=C89)N90C=CC(=C90)N91C=CC(=C91)N92C=CC(=C92)N93C=CC(=C93)N94C=CC(=C94)N95C=CC(=C95)N96C=CC(=C96)N97C=CC(=C97)N98C=CC(=C98)N99C=CC(=C99)N100C=CC(=C100)N101C=CC(=C101)N102C=CC(=C102)N103C=CC(=C103)N104C=CC(=C104)N105C=CC(=C105)N106C=CC(=C106)N107C=CC(=C107)N108C=CC(=C108)N109C=CC(=C109)N110C=CC(=C110)N111C=CC(=C111)N112C=CC(=C112)N113C=CC(=C113)N114C=CC(=C114)N115C=CC(=C115)N116C=CC(=C116)N117C=CC(=C117)N118C=CC(=C118)N119C=CC(=C119)N120C=CC(=C120)N121C=CC(=C121)N122C=CC(=C122)N123C=CC(=C123)N124C=CC(=C124)N125C=CC(=C125)N126C=CC(=C126)N127C=CC(=C127)N128C=CC(=C128)N129C=CC(=C129)N130C=CC(=C130)N131C=CC(=C131)N132C=CC(=C132)N133C=CC(=C133)N134C=CC(=C134)N135C=CC(=C135)N136C=CC(=C136)N137C=CC(=C137)N138C=CC(=C138)N139C=CC(=C139)N140C=CC(=C140)N141C=CC(=C141)N142C=CC(=C142)N143C=CC(=C143)N144C=CC(=C144)N145C=CC(=C145)N146C=CC(=C146)N147C=CC(=C147)N148C=CC(=C148)N149C=CC(=C149)N150C=CC(=C150)N151C=CC(=C151)N152C=CC(=C152)N153C=CC(=C153)N154C=CC(=C154)N155C=CC(=C155)N156C=CC(=C156)N157C=CC(=C157)N158C=CC(=C158)N159C=CC(=C159)N160C=CC(=C160)N161C=CC(=C161)N162C=CC(=C162)N163C=CC(=C163)N164C=CC(=C164)N165C=CC(=C165)N166C=CC(=C166)N167C=CC(=C167)N168C=CC(=C168)N169C=CC(=C169)N170C=CC(=C170)N171C=CC(=C171)N172C=CC(=C172)N173C=CC(=C173)N174C=CC(=C174)N175C=CC(=C175)N176C=CC(=C176)N177C=CC(=C177)N178C=CC(=C178)N179C=CC(=C179)N180C=CC(=C180)N181C=CC(=C181)N182C=CC(=C182)N183C=CC(=C183)N184C=CC(=C184)N185C=CC(=C185)N186C=CC(=C186)N187C=CC(=C187)N188C=CC(=C188)N189C=CC(=C189)N190C=CC(=C190)N191C=CC(=C191)N192C=CC(=C192)N193C=CC(=C193)N194C=CC(=C194)N195C=CC(=C195)N196C=CC(=C196)N197C=CC(=C197)N198C=CC(=C198)N199C=CC(=C199)N200C=CC(=C200)N201C=CC(=C201)N202C=CC(=C202)N203C=CC(=C203)N204C=CC(=C204)N205C=CC(=C205)N206C=CC(=C206)N207C=CC(=C207)N208C=CC(=C208)N209C=CC(=C209)N210C=CC(=C210)N211C=CC(=C211)N212C=CC(=C212)N213C=CC(=C213)N214C=CC(=C214)N215C=CC(=C215)N216C=CC(=C216)N217C=CC(=C217)N218C=CC(=C218)N219C=CC(=C219)N220C=CC(=C220)N221C=CC(=C221)N222C=CC(=C222)N223C=CC(=C223)N224C=CC(=C224)N225C=CC(=C225)N226C=CC(=C226)N227C=CC(=C227)N228C=CC(=C228)N229C=CC(=C229)N230C=CC(=C230)N231C=CC(=C231)N232C=CC(=C232)N233C=CC(=C233)N234C=CC(=C234)N235C=CC(=C235)N236C=CC(=C236)N237C=CC(=C237)N238C=CC(=C238)N239C=CC(=C239)N240C=CC(=C240)N241C=CC(=C241)N242C=CC(=C242)N243C=CC(=C243)N244C=CC(=C244)N245C=CC(=C245)N246C=CC(=C246)N247C=CC(=C247)N248C=CC(=C248)N249C=CC(=C249)N250C=CC(=C250)N251C=CC(=C251)N252C=CC(=C252)N253C=CC(=C253)N254C=CC(=C254)N255C=CC(=C255)N256C=CC(=C256)N257C=CC(=C257)N258C=CC(=C258)N259C=CC(=C259)N260C=CC(=C260)N261C=CC(=C261)N262C=CC(=C262)N263C=CC(=C263)N264C=CC(=C264)N265C=CC(=C265)N266C=CC(=C266)N267C=CC(=C267)N268C=CC(=C268)N269C=CC(=C269)N270C=CC(=C270)N271C=CC(=C271)N272C=CC(=C272)N273C=CC(=C273)N274C=CC(=C274)N275C=CC(=C275)N276C=CC(=C276)N277C=CC(=C277)N278C=CC(=C278)N279C=CC(=C279)N280C=CC(=C280)N281C=CC(=C281)N282C=CC(=C282)N283C=CC(=C283)N284C=CC(=C284)N285C=CC(=C285)N286C=CC(=C286)N287C=CC(=C287)N288C=CC(=C288)N289C=CC(=C289)N290C=CC(=C290)N291C=CC(=C291)N292C=CC(=C292)N293C=CC(=C293)N294C=CC(=C294)N295C=CC(=C295)N296C=CC(=C296)N297C=CC(=C297)N298C=CC(=C298)N299C=CC(=C299)N300C=CC(=C300)N301C=CC(=C301)N302C=CC(=C302)N303C=CC(=C303)N304C=CC(=C304)N305C=CC(=C305)N306C=CC(=C306)N307C=CC(=C307)N308C=CC(=C308)N309C=CC(=C309)N310C=CC(=C310)N311C=CC(=C311)N312C=CC(=C312)N313C=CC(=C313)N314C=CC(=C314)N315C=CC(=C315)N316C=CC(=C316)N317C=CC(=C317)N318C=CC(=C318)N319C=CC(=C319)N320C=CC(=C320)N321C=CC(=C321)N322C=CC(=C322)N323C=CC(=C323)N324C=CC(=C324)N325C=CC(=C325)N326C=CC(=C326)N327C=CC(=C327)N328C=CC(=C328)N329C=CC(=C329)N330C=CC(=C330)N331C=CC(=C331)N332C=CC(=C332)N333C=CC(=C333)N334C=CC(=C334)N335C=CC(=C335)N336C=CC(=C336)N337C=CC(=C337)N338C=CC(=C338)N339C=CC(=C339)N340C=CC(=C340)N341C=CC(=C341)N342C=CC(=C342)N343C=CC(=C343)N344C=CC(=C344)N345C=CC(=C345)N346C=CC(=C346)N347C=CC(=C347)N348C=CC(=C348)N349C=CC(=C349)N350C=CC(=C350)N351C=CC(=C351)N352C=CC(=C352)N353C=CC(=C353)N354C=CC(=C354)N355C=CC(=C355)N356C=CC(=C356)N357C=CC(=C357)N358C=CC(=C358)N359C=CC(=C359)N360C=CC(=C360)N361C=CC(=C361)N362C=CC(=C362)N363C=CC(=C363)N364C=CC(=C364)N365C=CC(=C365)N366C=CC(=C366)N367C=CC(=C367)N368C=CC(=C368)N369C=CC(=C369)N370C=CC(=C370)N371C=CC(=C371)N372C=CC(=C372)N373C=CC(=C373)N374C=CC(=C374)N375C=CC(=C375)N376C=CC(=C376)N377C=CC(=C377)N378C=CC(=C378)N379C=CC(=C379)N380C=CC(=C380)N381C=CC(=C381)N382C=CC(=C382)N383C=CC(=C383)N384C=CC(=C384)N385C=CC(=C385)N386C=CC(=C386)N387C=CC(=C387)N388C=CC(=C388)N389C=CC(=C389)N390C=CC(=C390)N391C=CC(=C391)N392C=CC(=C392)N393C=CC(=C393)N394C=CC(=C394)N395C=CC(=C395)N396C=CC(=C396)N397C=CC(=C397)N398C=CC(=C398)N399C=CC(=C399)N400C=CC(=C400)N401C=CC(=C401)N402C=CC(=C402)N403C=CC(=C403)N404C=CC(=C404)N405C=CC(=C405)N406C=CC(=C406)N407C=CC(=C407)N408C=CC(=C408)N409C=CC(=C409)N410C=CC(=C410)N411C=CC(=C411)N412C=CC(=C412)N413C=CC(=C413)N414C=CC(=C414)N415C=CC(=C415)N416C=CC(=C416)N417C=CC(=C417)N418C=CC(=C418)N419C=CC(=C419)N420C=CC(=C420)N421C=CC(=C421)N422C=CC(=C422)N423C=CC(=C423)N424C=CC(=C424)N425C=CC(=C425)N426C=CC(=C426)N427C=CC(=C427)N428C=CC(=C428)N429C=CC(=C429)N430C=CC(=C430)N431C=CC(=C431)N432C=CC(=C432)N433C=CC(=C433)N434C=CC(=C434)N435C=CC(=C435)N436C=CC(=C436)N437C=CC(=C437)N438C=CC(=C438)N439C=CC(=C439)N440C=CC(=C440)N441C=CC(=C441)N442C=CC(=C442)N443C=CC(=C443)N444C=CC(=C444)N445C=CC(=C445)N446C=CC(=C446)N447C=CC(=C447)N448C=CC(=C448)N449C=CC(=C449)N450C=CC(=C450)N451C=CC(=C451)N452C=CC(=C452)N453C=CC(=C453)N454C=CC(=C454)N455C=CC(=C455)N456C=CC(=C456)N457C=CC(=C457)N458C=CC(=C458)N459C=CC(=C459)N460C=CC(=C460)N461C=CC(=C461)N462C=CC(=C462)N463C=CC(=C463)N464C=CC(=C464)N465C=CC(=C465)N466C=CC(=C466)N467C=CC(=C467)N468C=CC(=C468)N469C=CC(=C469)N470C=CC(=C470)N471C=CC(=C471)N472C=CC(=C472)N473C=CC(=C473)N474C=CC(=C474)N475C=CC(=C475)N476C=CC(=C476)N477C=CC(=C477)N478C=CC(=C478)N479C=CC(=C479)N480C=CC(=C480)N481C=CC(=C481)N482C=CC(=C482)N483C=CC(=C483)N484C=CC(=C484)N485C=CC(=C485)N486C=CC(=C486)N487C=CC(=C487)N488C=CC(=C488)N489C=CC(=C489)N490C=CC(=C490)N491C=CC(=C491)N492C=CC(=C492)N493C=CC(=C493)N494C=CC(=C494)N495C=CC(=C495)N496C=CC(=C496)N497C=CC(=C497)N498C=CC(=C498)N499C=CC(=C499)N500C=CC(=C500)N501C=CC(=C501)N502C=CC(=C502)N503C=CC(=C503)N504C=CC(=C504)N505C=CC(=C505)N506C=CC(=C506)N507C=CC(=C507)N508C=CC(=C508)N509C=CC(=C509)N510C=CC(=C510)N511C=CC(=C511)N512C=CC(=C512)N513C=CC(=C513)N514C=CC(=C514)N515C=CC(=C515)N516C=CC(=C516)N517C=CC(=C517)N518C=CC(=C518)N519C=CC(=C519)N520C=CC(=C520)N521C=CC(=C521)N522C=CC(=C522)N523C=CC(=C523)N524C=CC(=C524)N525C=CC(=C525)N526C=CC(=C526)N527C=CC(=C527)N528C=CC(=C528)N529C=CC(=C529)N530C=CC(=C530)N531C=CC(=C531)N532C=CC(=C532)N533C=CC(=C533)N534C=CC(=C534)N535C=CC(=C535)N536C=CC(=C536)N537C=CC(=C537)N538C=CC(=C538)N539C=CC(=C539)N540C=CC(=C540)N541C=CC(=C541)N542C=CC(=C542)N543C=CC(=C543)N544C=CC(=C544)N545C=CC(=C545)N546C=CC(=C546)N547C=CC(=C547)N548C=CC(=C548)N549C=CC(=C549)N550C=CC(=C550)N551C=CC(=C551)N552C=CC(=C552)N553C=CC(=C553)N554C=CC(=C554)N555C=CC(=C555)N556C=CC(=C556)N557C=CC(=C557)N558C=CC(=C558)N559C=CC(=C559)N560C=CC(=C560)N561C=CC(=C561)N562C=CC(=C562)N563C=CC(=C563)N564C=CC(=C564)N565C=CC(=C565)N566C=CC(=C566)N567C=CC(=C567)N568C=CC(=C568)N569C=CC(=C569)N570C=CC(=C570)N571C=CC(=C571)N572C=CC(=C572)N573C=CC(=C573)N574C=CC(=C574)N575C=CC(=C575)N576C=CC(=C576)N577C=CC(=C577)N578C=CC(=C578)N579C=CC(=C579)N580C=CC(=C580)N581C=CC(=C581)N582C=CC(=C582)N583C=CC(=C583)N584C=CC(=C584)N585C=CC(=C585)N586C=CC(=C586)N587C=CC(=C587)N588C=CC(=C588)N589C=CC(=C589)N590C=CC(=C590)N591C=CC(=C591)N592C=CC(=C592)N593C=CC(=C593)N594C=CC(=C594)N595C=CC(=C595)N596C=CC(=C596)N597C=CC(=C597)N598C=CC(=C598)N599C=CC(=C599)N600C=CC(=C600)N601C=CC(=C601)N602C=CC(=C602)N603C=CC(=C603)N604C=CC(=C604)N605C=CC(=C605)N606C=CC(=C606)N607C=CC(=C607)N608C=CC(=C608)N609C=CC(=C609)N610C=CC(=C610)N611C=CC(=C611)N612C=CC(=C612)N613C=CC(=C613)N614C=CC(=C614)N615C=CC(=C615)N616C=CC(=C616)N617C=CC(=C617)N618C=CC(=C618)N619C=CC(=C619)N620C=CC(=C620)N621C=CC(=C621)N622C=CC(=C622)N623C=CC(=C623)N624C=CC(=C624)N625C=CC(=C625)N626C=CC(=C626)N627C=CC(=C627)N628C=CC(=C628)N629C=CC(=C629)N630C=CC(=C630)N631C=CC(=C631)N632C=CC(=C632)N633C=CC(=C633)N634C=CC(=C634)N635C=CC(=C635)N636C=CC(=C636)N637C=CC(=C637)N638C=CC(=C638)N639C=CC(=C639)N640C=CC(=C640)N641C=CC(=C641)N642C=CC(=C642)N643C=CC(=C643)N644C=CC(=C644)N645C=CC(=C645)N646C=CC(=C646)N647C=CC(=C647)N648C=CC(=C648)N649C=CC(=C649)N650C=CC(=C650)N651C=CC(=C651)N652C=CC(=C652)N653C=CC(=C653)N654C=CC(=C654)N655C=CC(=C655)N656C=CC(=C656)N657C=CC(=C657)N658C=CC(=C658)N659C=CC(=C659)N660C=CC(=C660)N661C=CC(=C661)N662C=CC(=C662)N663C=CC(=C663)N664C=CC(=C664)N665C=CC(=C665)N666C=CC(=C666)N667C=CC(=C667)N668C=CC(=C668)N669C=CC(=C669)N670C=CC(=C670)N671C=CC(=C671)N672C=CC(=C672)N673C=CC(=C673)N674C=CC(=C674)N675C=CC(=C675)N676C=CC(=C676)N677C=CC(=C677)N678C=CC(=C678)N679C=CC(=C679)N680C=CC(=C680)N681C=CC(=C681)N682C=CC(=C682)N683C=CC(=C683)N684C=CC(=C684)N685C=CC(=C685)N686C=CC(=C686)N687C=CC(=C687)N688C=CC(=C688)N689C=CC(=C689)N690C=CC(=C690)N691C=CC(=C691)N692C=CC(=C692)N693C=CC(=C693)N694C=CC(=C694)N695C=CC(=C695)N696C=CC(=C696)N697C=CC(=C697)N698C=CC(=C698)N699C=CC(=C699)N700C=CC(=C700)N701C=CC(=C701)N702C=CC(=C702)N703C=CC(=C703)N704C=CC(=C704)N705C=CC(=C705)N706C=CC(=C706)N707C=CC(=C707)N708C=CC(=C708)N709C=CC(=C709)N710C=CC(=C710)N711C=CC(=C711)N712C=CC(=C712)N713C=CC(=C713)N714C=CC(=C714)N715C=CC(=C715)N716C=CC(=C716)N717C=CC(=C717)N718C=CC(=C718)N719C=CC(=C719)N720C=CC(=C720)N721C=CC(=C721)N722C=CC(=C722)N723C=CC(=C723)N724C=CC(=C724)N725C=CC(=C725)N726C=CC(=C726)N727C=CC(=C727)N728C=CC(=C728)N729C=CC(=C729)N730C=CC(=C730)N731C=CC(=C731)N732C=CC(=C732)N733C=CC(=C733)N734C=CC(=C734)N735C=CC(=C735)N736C=CC(=C736)N737C=CC(=C737)N738C=CC(=C738)N739C=CC(=C739)N740C=CC(=C740)N741C=CC(=C741)N742C=CC(=C742)N743C=CC(=C743)N744C=CC(=C744)N745C=CC(=C745)N746C=CC(=C746)N747C=CC(=C747)N748C=CC(=C748)N749C=CC(=C749)N750C=CC(=C750)N751C=CC(=C751)N752C=CC(=C752)N753C=CC(=C753)N754C=CC(=C754)N755C=CC(=C755)N756C=CC(=C756)N757C=CC(=C757)N758C=CC(=C758)N759C=CC(=C759)N760C=CC(=C760)N761C=CC(=C761)N762C=CC(=C762)N763C=CC(=C763)N764C=CC(=C764)N765C=CC(=C765)N766C=CC(=C766)N767C=CC(=C767)N768C=CC(=C768)N769C=CC(=C769)N770C=CC(=C770)N771C=CC(=C771)N772C=CC(=C772)N773C=CC(=C773)N774C=CC(=C774)N775C=CC(=C775)N776C=CC(=C776)N777C=CC(=C777)N778C=CC(=C778)N779C=CC(=C779)N780C=CC(=C780)N781C=CC(=C781)N782C=CC(=C782)N783C=CC(=C783)N784C=CC(=C784)N785C=CC(=C785)N786C=CC(=C786)N787C=CC(=C787)N788C=CC(=C788)N789C=CC(=C789)N790C=CC(=C790)N791C=CC(=C791)N792C=CC(=C792)N793C=CC(=C793)N794C=CC(=C794)N795C=CC(=C795)N796C=CC(=C796)N797C=CC(=C797)N798C=CC(=C798)N799C=CC(=C799)N800C=CC(=C800)N801C=CC(=C801)N802C=CC(=C802)N803C=CC(=C803)N804C=CC(=C804)N805C=CC(=C805)N806C=CC(=C806)N807C=CC(=C807)N808C=CC(=C808)N809C=CC(=C809)N810C=CC(=C810)N811C=CC(=C811)N812C=CC(=C812)N813C=CC(=C813)N814C=CC(=C814)N815C=CC(=C815)N816C=CC(=C816)N817C=CC(=C817)N818C=CC(=C818)N819C=CC(=C819)N820C=CC(=C820)N821C=CC(=C821)N822C=CC(=C822)N823C=CC(=C823)N824C=CC(=C824)N825C=CC(=C825)N826C=CC(=C826)N827C=CC(=C827)N828C=CC(=C828)N829C=CC(=C829)N830C=CC(=C830)N831C=CC(=C831)N832C=CC(=C832)N833C=CC(=C833)N834C=CC(=C834)N835C=CC(=C835)N836C=CC(=C836)N837C=CC(=C837)N838C=CC(=C838)N839C=CC(=C839)N840C=CC(=C840)N841C=CC(=C841)N842C=CC(=C842)N843C=CC(=C843)N844C=CC(=C844)N845C=CC(=C845)N846C=CC(=C846)N847C=CC(=C847)N848C=CC(=C848)N849C=CC(=C849)N850C=CC(=C850)N851C=CC(=C851)N852C=CC(=C852)N853C=CC(=C853)N854C=CC(=C854)N855C=CC(=C855)N856C=CC(=C856)N857C=CC(=C857)N858C=CC(=C858)N859C=CC(=C859)N860C=CC(=C860)N861C=CC(=C861)N862C=CC(=C862)N863C=CC(=C863)N864C=CC(=C864)N865C=CC(=C865)N866C=CC(=C866)N867C=CC(=C867)N868C=CC(=C868)N869C=CC(=C869)N870C=CC(=C870)N871C=CC(=C871)N872C=CC(=C872)N87</chem>			

# Partner Extensions at Work: Spotfire



# Partner Extensions at Work: Spotfire



The image displays the KNIME software interface, illustrating the integration of Spotfire for data visualization. The top window shows a workflow with five nodes: File Reader (Node 1), Color Manager (Node 2), Open TIBCO Spotfire (Node 3), TIBCO File Writer (Node 4), and Scatter Plot (Node 5). The Scatter Plot node is currently active, displaying a scatter plot of data points colored by class (Iris-setosa, Iris-versicolor, Iris-virginica).

Below the workflow, two Spotfire windows are shown, each displaying a scatter plot of the same data. The left window shows the plot with a grid and axes labeled 'sepal width' and 'petal width'. The right window shows the plot with a grid and axes labeled 'sepal width' and 'petal width'. Both plots have a legend on the right side, indicating the color coding for the three Iris species: Iris-setosa (red), Iris-versicolor (green), and Iris-virginica (blue).

The Spotfire interface includes a menu bar (File, Edit, View, Insert, Tools, Help) and a toolbar with various icons for data manipulation and visualization. The status bar at the bottom of each Spotfire window indicates 'Online', '150 of 150 rows', '50 marked', and '6 columns'.

„I didn‘t mean someone else‘s tool -  
our own is so much better.

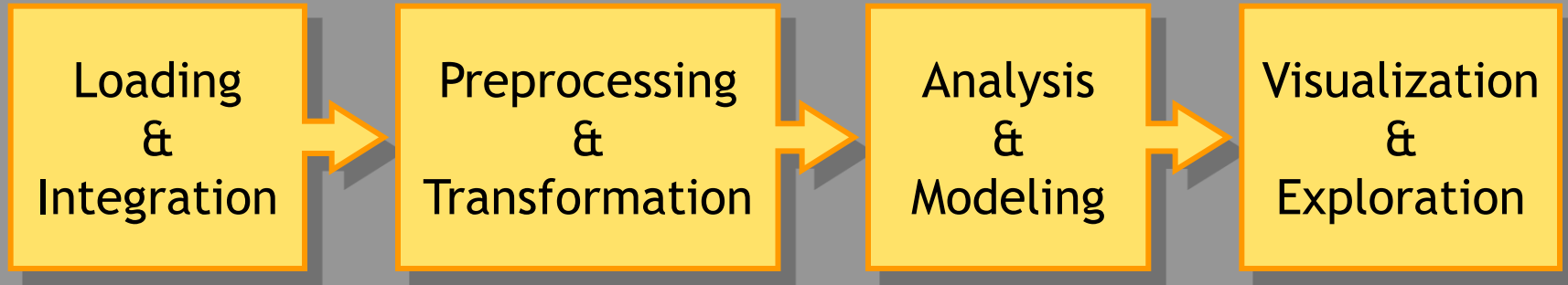
(nobody understands it anymore but everybody uses it)!“



# Integration of Tools, cont.

- External Tool Node  
(launches shell commands)
- Web Service node
- New Node Wizard
  - for thin wrappers
  - for your own prototyping / development!

## “Integrating Data, Tools, and Science”



**Data:** Integrate distributed & heterogeneous data sources

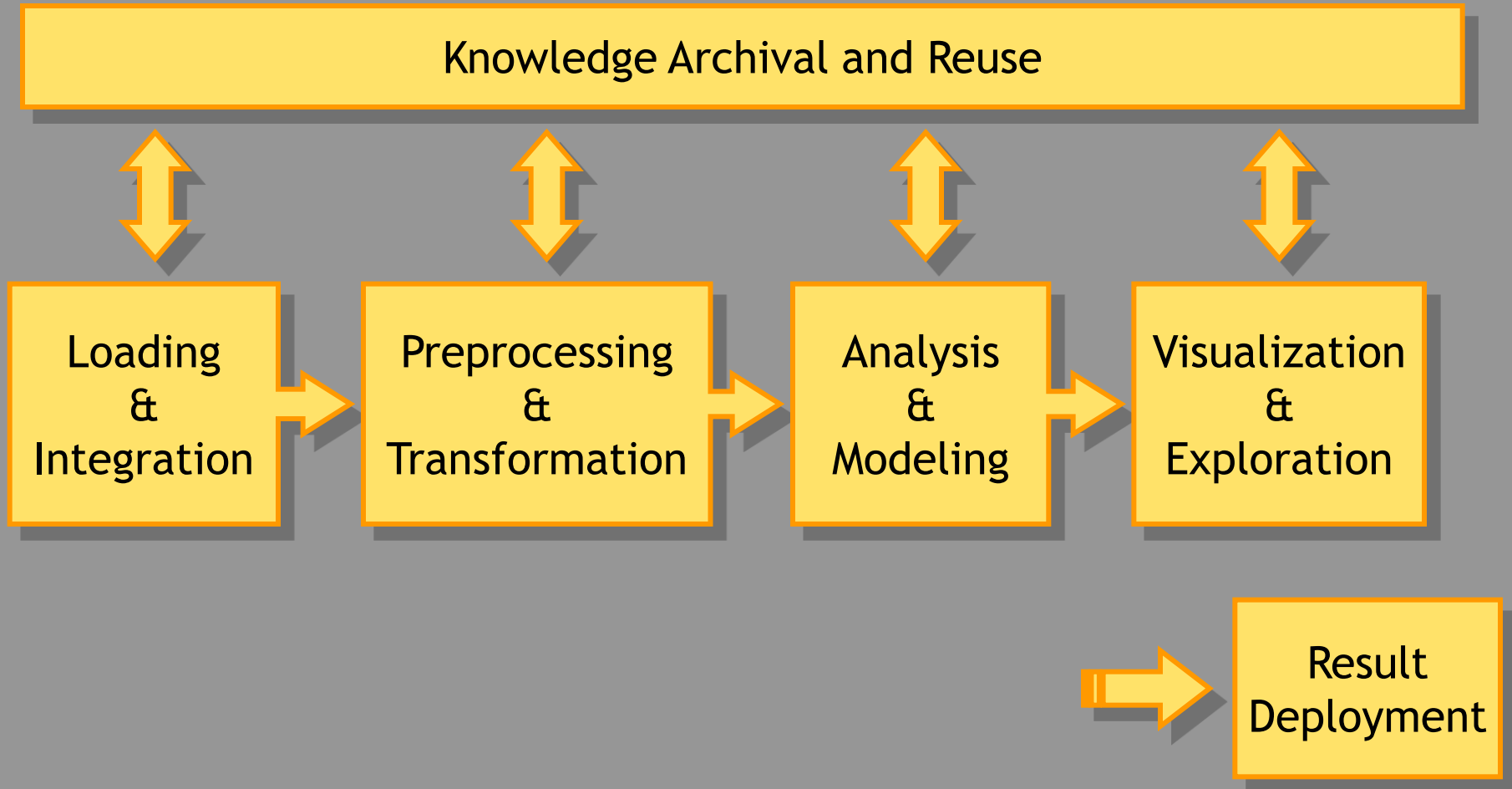
**Tools:** Integrate Open Source, Third Party and In-house Tools

**Science:** Standardize Analysis and Modeling Workflows

# „Very nice, but...“

- „...I want to share workflows!“
- „...I want to run workflows on a server...“
- „...and our cluster!“
- „...I want to show the results to upper management!“

# The Larger Picture



**Wait a second - is this still research?**

**June 2008:**

**KNIME.com GmbH  
founded in Zurich, Switzerland**

# KNIME.com GmbH

## Software & Services:

- Support for Open Source KNIME Platform
- Consulting
- Development:
  - Customizations
  - KNIME Extensions
  - Enterprise Components

## Knowledge Archival and Reuse

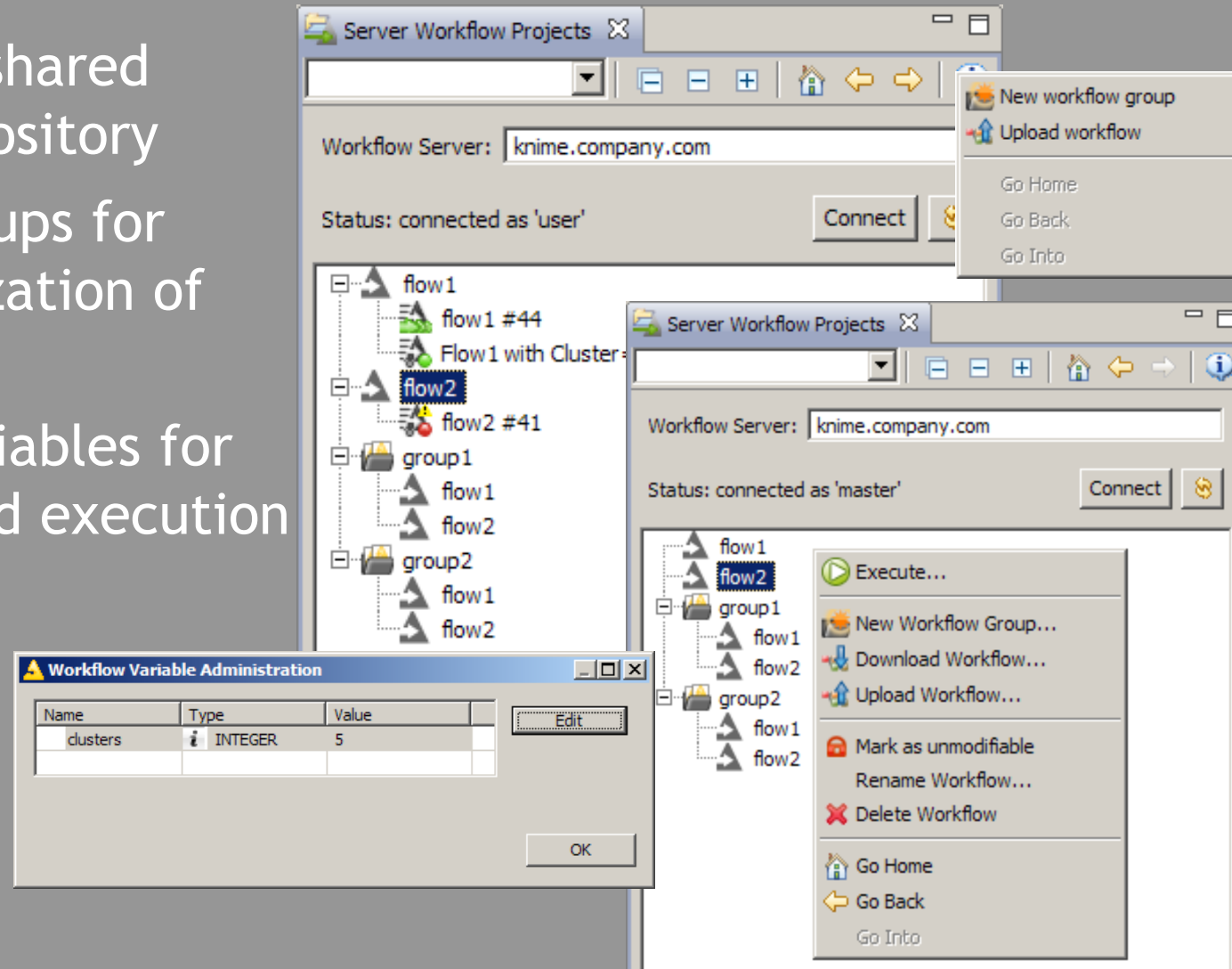
- Workflow Repository:
    - predefined flows: standardization
    - Workflow archival: knowledge repository
  - Metanode Repository:
    - workflow fragments: expert knowledge archive
- ⇒ KNIME Enterprise Server
- Central corporate repository for workflows
  - User access control
  - Scheduled execution
  - Backend data processing engine for SOA integration



# Enterprise Server

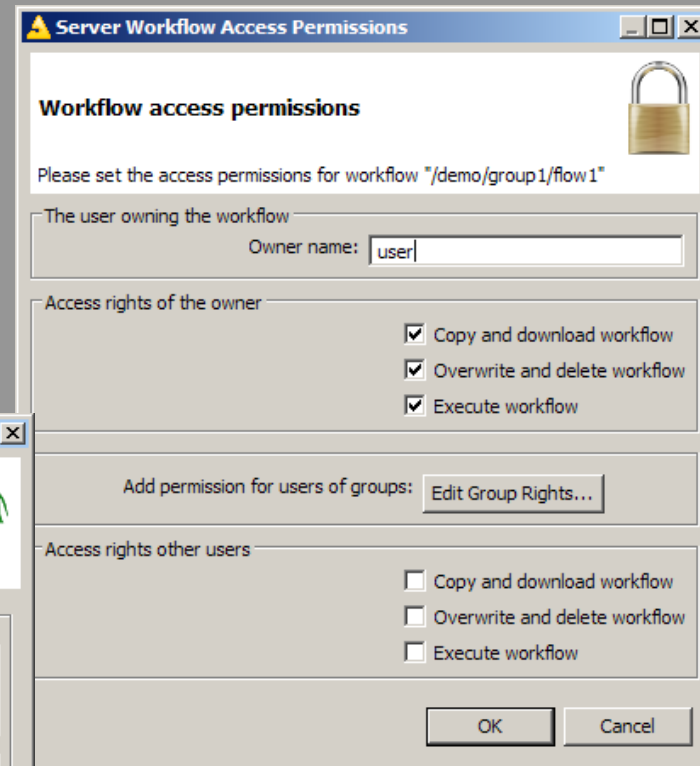
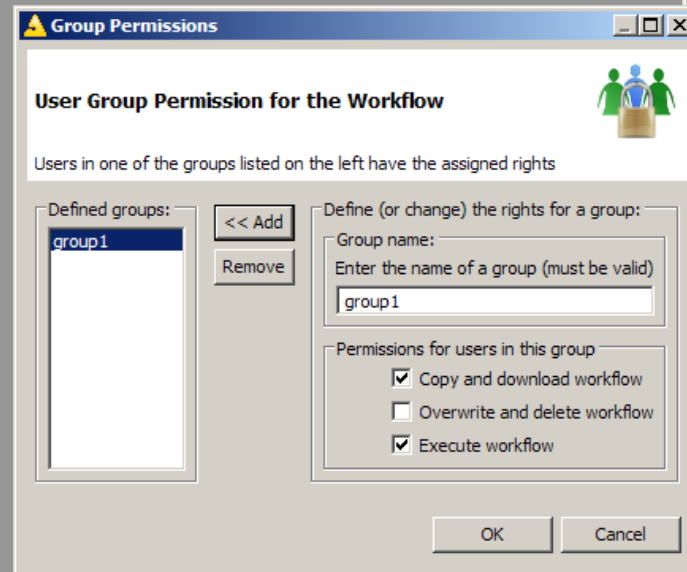


- Centralized, shared workflow repository
- Workflow groups for easier organization of repository
- Workflow Variables for parameterized execution



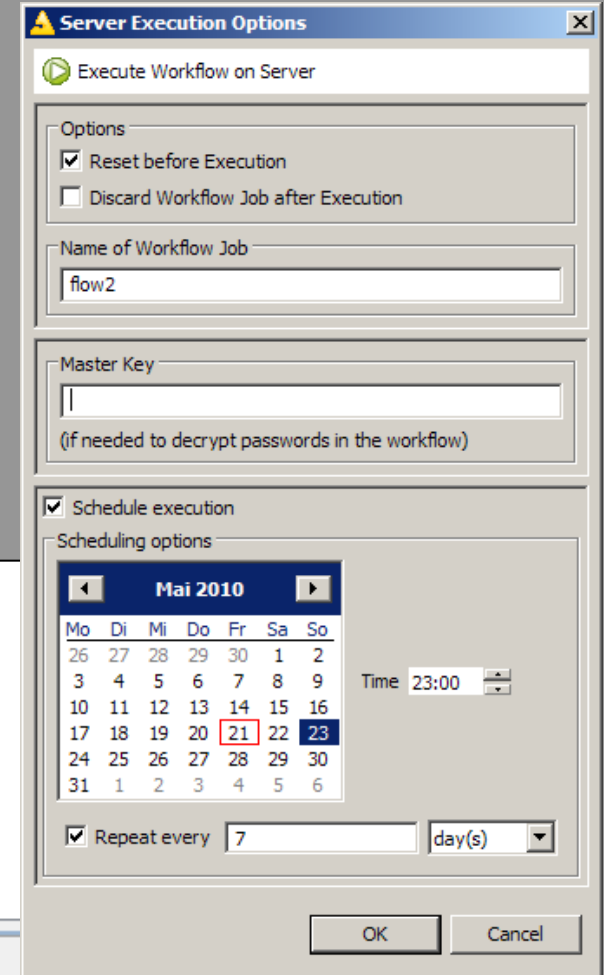
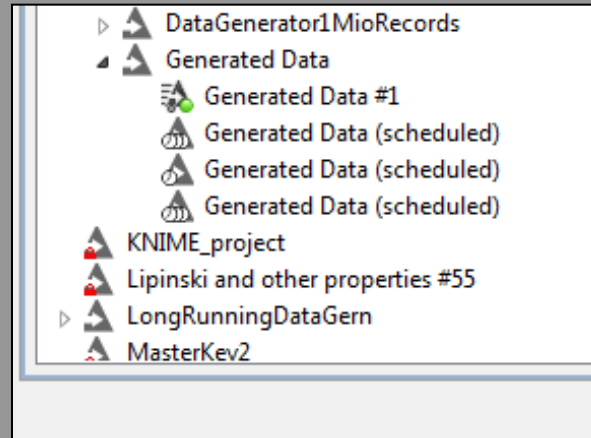
# User Rights Management

- Controlled, fine granular access to workflows
- Connects to existing user authentication setup



# Scheduled Execution

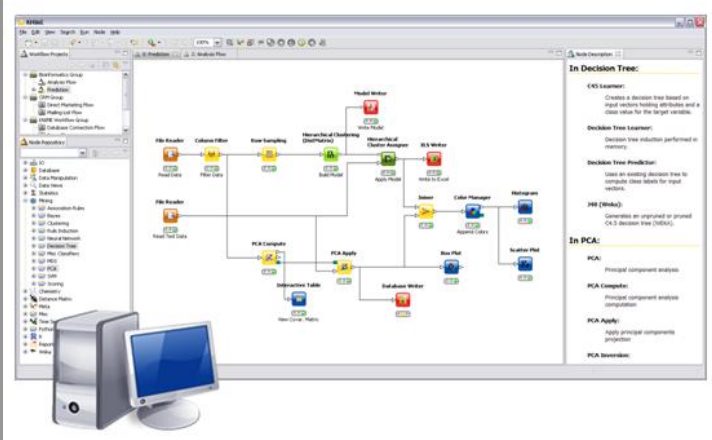
- Manual server side execution...
- ...or scheduled for automated, regular execution.



## Result Deployment - 1

- Reports:
    - KNIME workflows load, aggregate, analyze data
    - Reports summarize and enrich workflow output (tables, graphs, images) - a BIRT integration, btw...
    - “one click” update for new data or parameters
- ⇒ KNIME Report Designer and Server
- Provides access to KNIME workflows without exposing the underlying workflows and nodes
  - Templates created by expert
  - End users execute template and export report to desired format (PPT, DOC, PDF, HTML, ...)

# KNIME Report Server



with KNIME Report Designer. Provided by KNIME.com GmbH, Zurich, Switzerland

## Result Deployment - 2

- Models:
    - Export of Trained Models
  - Workflows
    - Access to workflows from other applications
- ⇒ PMML Export of models
- ⇒ Access to workflows via web service calls

# Just another Data Analysis Workbench?

# KNIME: the backbone of your data



## Specialized Extensions

- Chemo and Bio Informatics
- Image Analysis
- Time Series, Text, ...
- ... Partner Tools

## Analytics / Modeling

- Built-In Data Analytics
- R-Project
- Weka, LibSVM, CDK...
- ...Partner Tools

## Exploration

- Interactive Views
- KNIME Reports
- ...Spotfire

## Integration

- KWaaS

## Information Backbone: KNIME Workflows

## Legacy SW

- Ext Tool
- WS / PP

## Collaboration

- KNIME Enterprise Server

## HPC

- Cluster Execution
- KNIME o.t. Cloud

## Deployment

- KNIME Report Server
- PMML
- KWaaS



# Status Quo

- KNIME 2.2.0 released (July 2010)
- >5,500 active users
  - ~50% Life Science
  - ~50% Business Intelligence, Analytics
- 10+ commercial partners
- Distributors in Japan, Italy, China, Brazil
- Since Aug 2010: Global Support and Distribution Agreement with CambridgeSoft

A blue badge with white text that reads "Gartner | 2010 COOL VENDOR".

**Gartner | 2010  
COOL VENDOR**

# Integration!

- **Data:** form one view for all activities (transformation, modeling, analytics, reporting)
- **Tools:** open source and commercial (own and others)
- **Science (&Expertise):** archive and reuse workflows

try it out - it's free:

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