

# How can Pipeline Pilot help you?

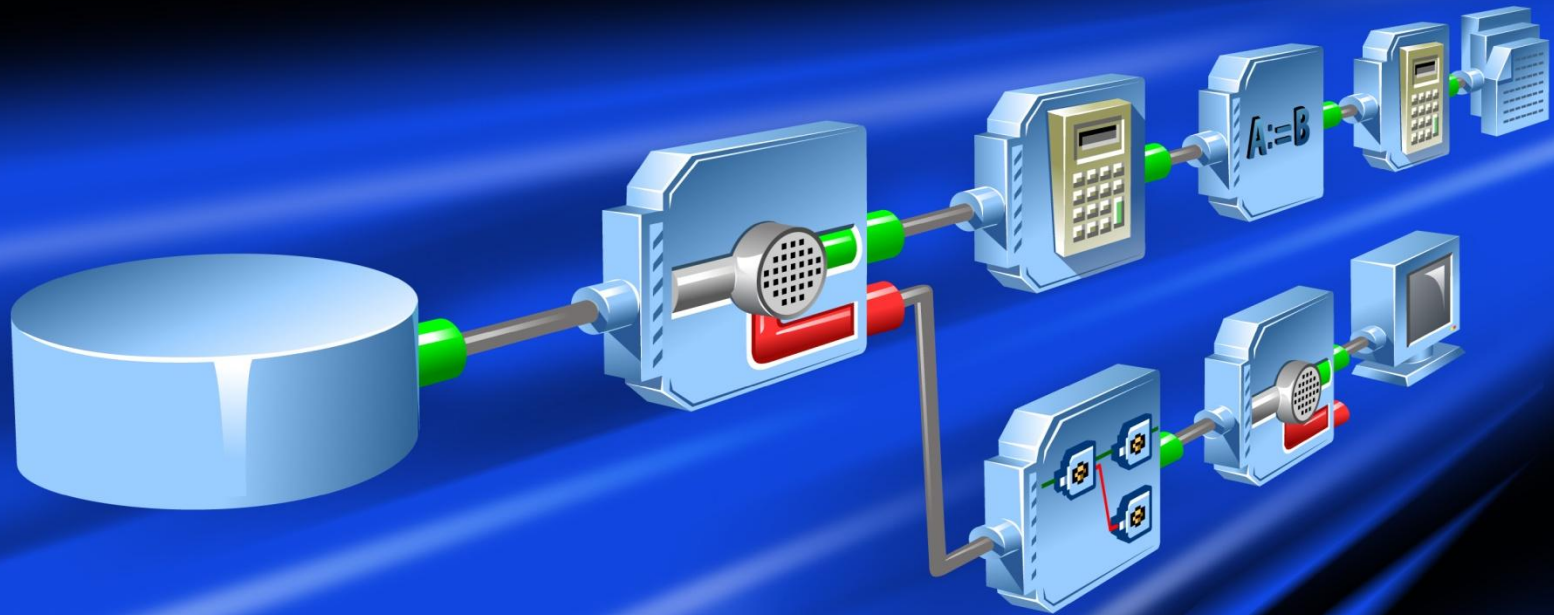
Semin-R 25 Nov. 2010

Stéphane Vellay – [SVellay@accelrys.com](mailto:SVellay@accelrys.com)

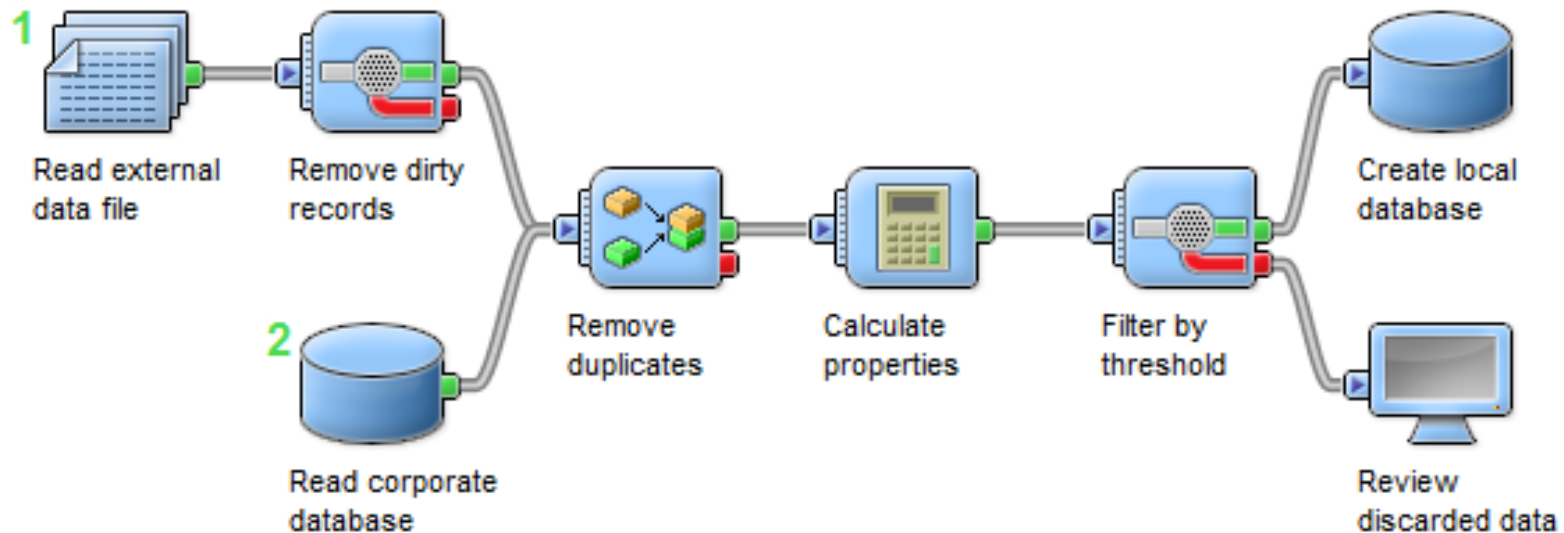


## Enterprise software for

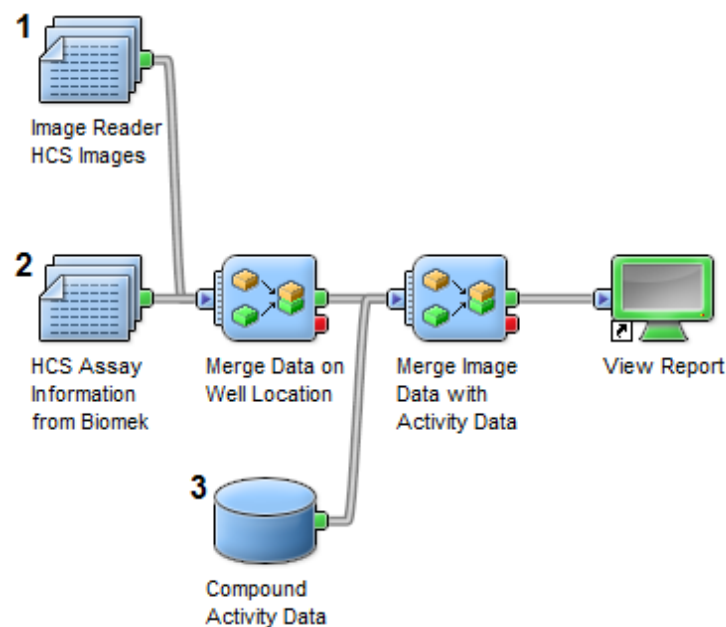
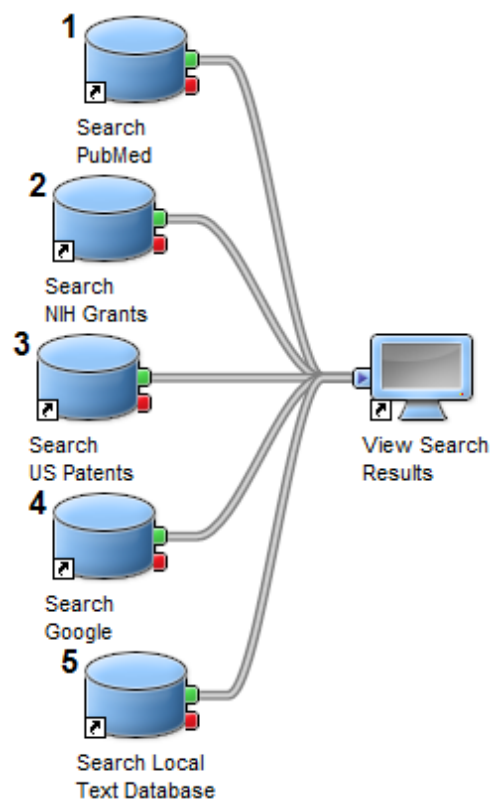
- Integrating diverse data & applications
- Automating data processing & analysis
- Capturing & deploying best-practice processes



- A powerful new paradigm for data processing
- Pipelines guide the flow of data through a network of modular computational components

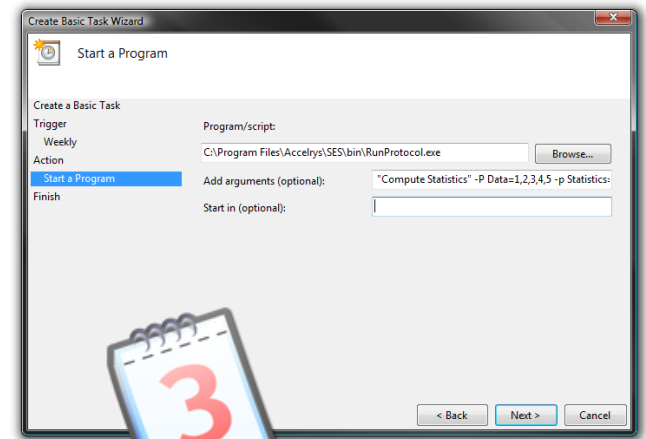
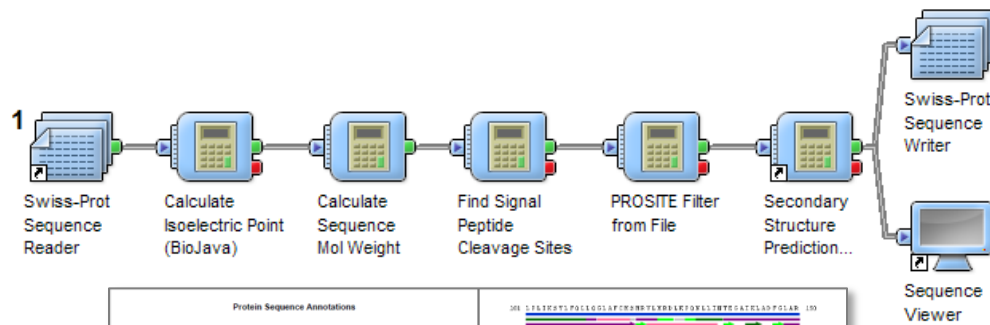


- Access disparate research data locked in silos
- Automate the scientific analysis
- Explore, visualize and report research results



# With Pipeline Pilot

- Access disparate research data locked in silos
- Automate the scientific analysis
- Explore, visualize and report research results



- Access disparate research data locked in silos
- Automate the scientific analysis
- Explore, visualize and report research results

Find and align substructures

Query structure: c1ccc(cc1)-c2ccccc2

Molecule: CC1=CC=C(C=C1)C2=CC=CC=C2

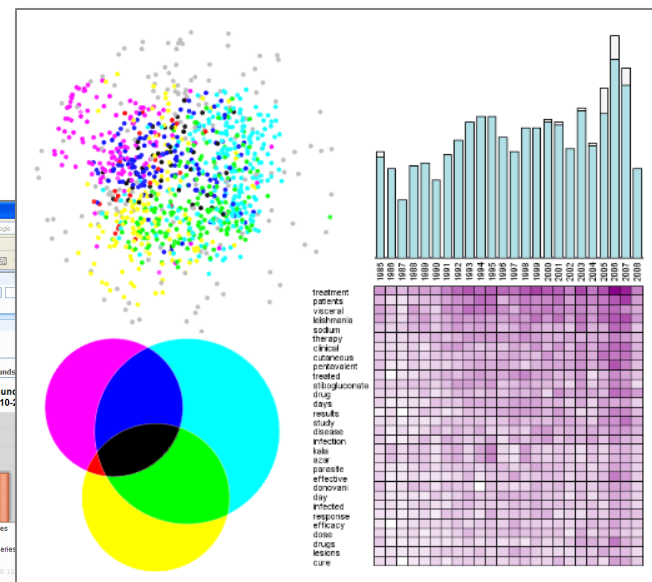
CODE: SPB 0002 | IUPAC NAME: 2-(5H-tetrahydro-3H-imidazol-5-yl)phenol

Database to Search:  Asinex,  Sample Drugs,  Other (Specify Below...)

Optional: Specify SD File:

Maximum compounds to search: 5000

Image	ID	Assay 1	Assay 2	Met Weight	Allogl	Prod Sol	Entry Date	
	PR024751	0.07	0.71	0.76	810.7208	2.32	-0.75	2008-08-01
	PR075437	0.38	0.31	0.76	402.48689	1.619	-3.91	2008-05-02
	PR012246	0.77	0.29	0.54	504.65246	-2.172	-1.23	2008-04-24
	PR022006	0.76	0.09	0.09	540.54299	5.85	-12.00	2008-06-01



## Gather Data

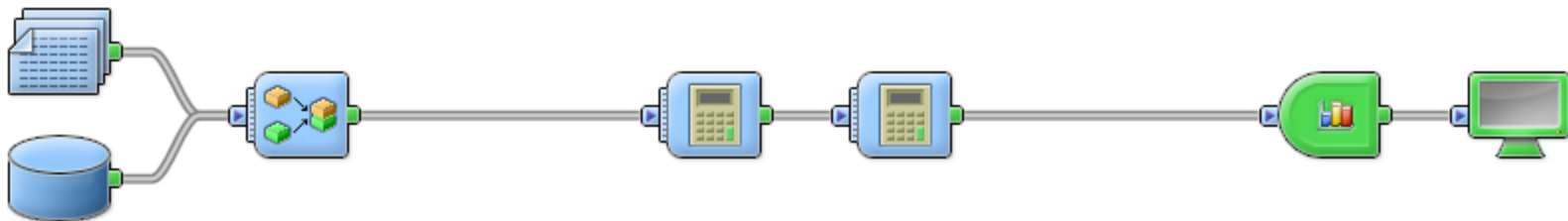
- Extract knowledge
- Fewer errors

## Automate Applications

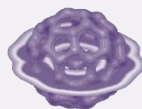
- Extract knowledge
- Fewer errors
- Save time
- Focus on Science

## Report & Deploy

- Extract knowledge
- Fewer errors
- Save time
- Focus on Science
- Educate decisions
- Collaborate



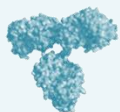
## Materials M&S



[Materials Studio](#)

[Polymer Properties \(Synthia\)](#)

## Life Science M&S



[Catalyst \(Pharmacophore\)](#)

[CHARMm \(Simulations\)](#)

## Chemistry



[Chemistry](#)

[ADMET](#)

[Cheminformatics](#)

## Biology



[Sequence Analysis](#)

[Gene Expression](#)

[Mass Spec. for Proteomics](#)

[Next Generation Sequencing](#)

## Laboratory



[Plate Data Analytics](#)

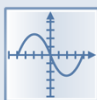
[Analytical Instrumentation](#)

## Imaging



[Imaging](#)

## Analysis & Statistics



[Data Modeling](#)

[Advanced Data Modeling](#)

[R Statistics](#)

## Document Search & Analysis



[Text Analytics](#)

[Chemical Text Mining](#)

## Database & Application Integration



[Integration](#)

[Client-Side SDKs](#)

## Reporting & Visualization



[Reporting](#)



# R Statistics Collection

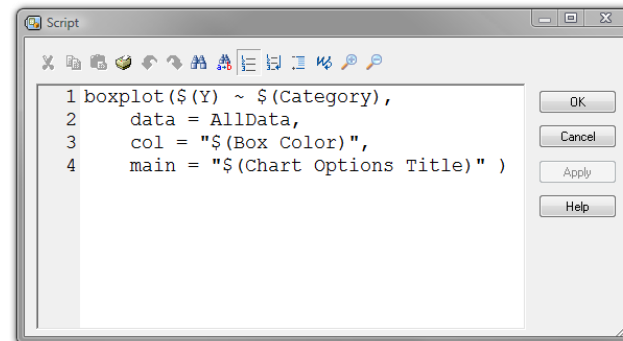
Data Pipelining for Discovery Informatics



- Facilitate easy integration of R in Pipeline Pilot workflow
  - Capitalize on existing knowledge in R and make it available to all scientist without additional training
  - Let scientists focus on science, not on a specific knowledge
- Out of the box R-based component for
  - Learning and apply models: SVM, neural network, etc.
  - Dataset analysis: clustering, ANOVA, PCA, etc.
  - Statistics specific visualisation: Conditional Plots, Pair plots, Box plots, etc.

## for Experts within Pipeline Pilot

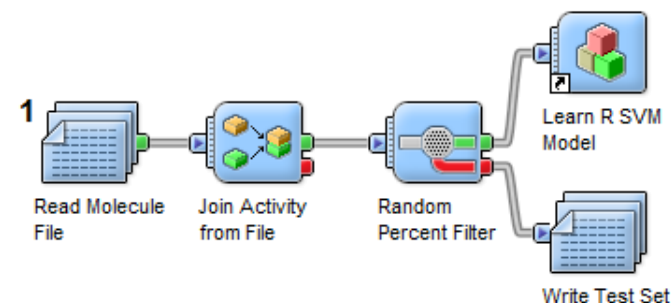
- **Familiar & extensible language**
- **Share** existing R scripts



```
1 boxplot($ (Y) ~ $ (Category),  
2 data = AllData,  
3 col = "$ (Box Color)",  
4 main = "$ (Chart Options Title)" )
```

## for Scientists without scripting

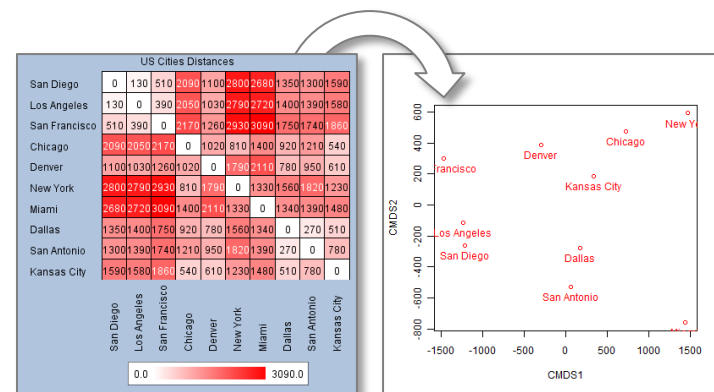
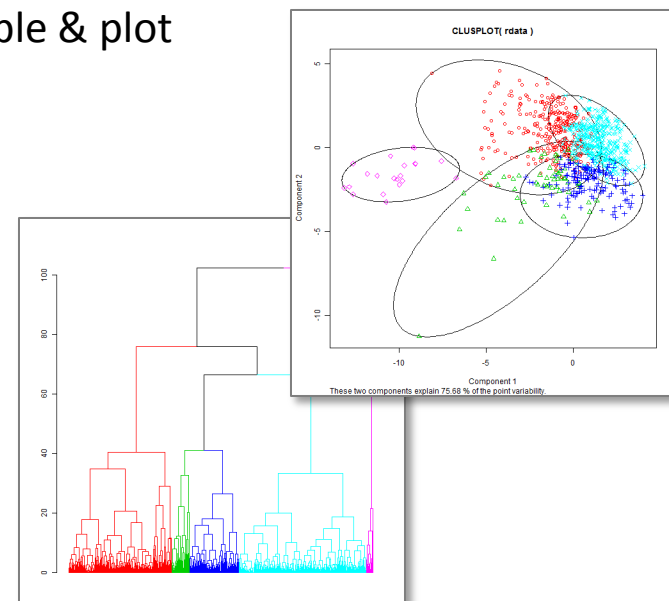
- **Reuse** existing R scripts
- **Focus** on Statistics & Science



## ○ Synergies

- Statistically-aware R & Scientifically-aware Pipeline Pilot
- Insightful analyses & educated decision

- Demo
  - Read Iris, R Correlation Matrix, report correlation table & plot
- R Graphics
  - Histograms and Pairs Plots
- R Clustering
  - Hierarchical Clustering of Asinex Molecules
  - Glycine Clustering (Analytical Instrumentation)
- R Learn Models
  - Free-Wilson QSAR Analysis with R Linear Model
- R Analysis
  - Principal Components Analysis of Asinex molecules
  - R Custom Script Example
- R Multidimensional Scaling
  - Location of US cities from distances
  - Visualization of 20 Aminoacids in MDS Space



- Access disparate research data locked in silos
  - Extract knowledge
  - Fewer errors
- Automate the scientific analysis
  - Save time
  - Focus on Science
- Explore, visualize and report research results
  - Educate decisions
  - Collaborate

# Questions & Comments

Accelrys Web Site

 [www.accelrys.com](http://www.accelrys.com)

Accelrys Community

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

Accelrys Advantage

 [customer.accelrys.com](http://customer.accelrys.com)

Accelrys Training

 [training@accelrys.com](mailto:training@accelrys.com)

Pipeline Pilot Support

 [support@accelrys.com](mailto:support@accelrys.com)

