aneurIST

Integrated biomedical informatics for the management of cerebral aneurysms

Workflows for Data Mining in integrated multi-modal data on Intracranial Aneurysms using KNIME

Roelof Risselada, Christoph M. Friedrich, Christian Ebeling, Roman Klinger, Anna Bauer-Mehren, Manuel Pastor, Maria Cruz Villa, Jose M. Pozo, Alejandro F. Frangi, and Martin Hofmann-Apitius

UseR! Conference, 8 July 2009, Rennes, France

The financial support of the European Commission is gratefully acknowledged. Material in this presentation reflects only the author’s views and the Commission is not liable for any use that may be made of the information contained herein.

Outline

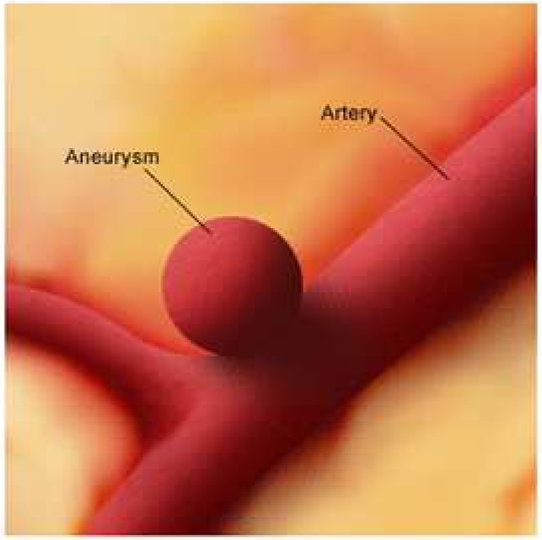
- Introduction
  - Model disease
  - @neurIST project
  - Problem
  - Route to solution
- Introductory to KNIME
- Demonstration KNIME in model disease
- R application in KNIME
- Obtained results – an example

Model disease

- Intracranial Aneurysm (IA): a bulbous expansion of a brain artery
- Prevalence of 2-5% in the EU population
- Risk of rupture leading to severe bleeding: 0.7 % per year
- Better imaging → more asymptomatic IA are detected
- Therapy: surgical procedure or *watchful waiting*
- Aim: predict outcome

- EU project: @neurIST

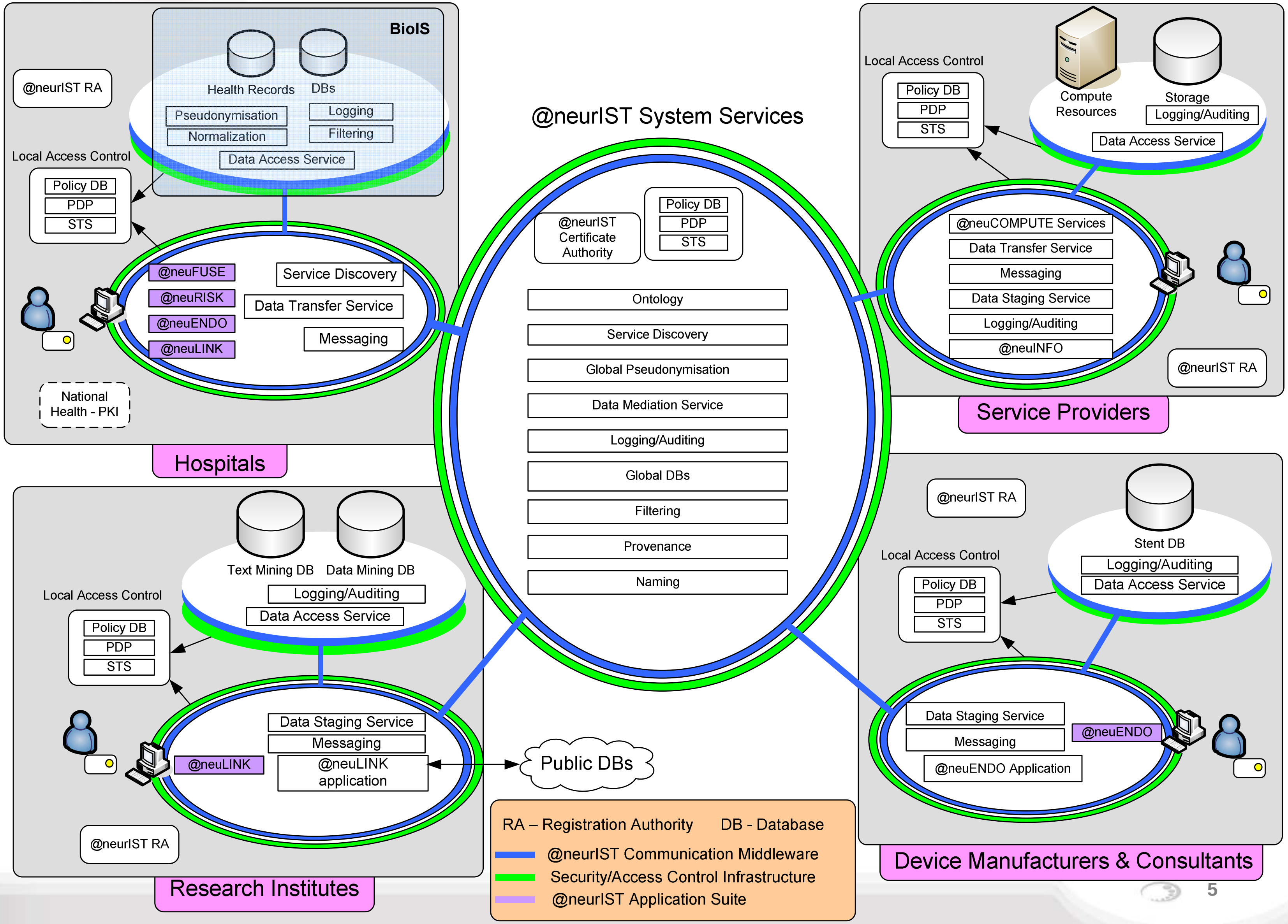
Cerebral Aneurysm



@neurIST

- Development of an integrated healthcare infrastructure to improve decision support for IA
- Integrated European FP6 healthcare GRID Project with 32 partners http://www.aneurist.org
- Objective: predict the risk of rupture of IA for an individual patient
- Multi-modal data
  - Imaging data: morphology, haemodynamic models
  - Clinical data
  - Genetic data (SNP)
  - Epidemiological data
  - Literature data

@neurIST System Services



@neurIST Certificate Authority
Policy DB
PDP
STS
Ontology
Service Discovery
Global Pseudonymisation
Data Mediation Service
Logging/Auditing
Global DBs
Filtering
Provenance
Naming

Hospitals

BioIS
Health Records
DBs
Pseudonymisation
Normalization
Logging
Filtering
Data Access Service
@neurIST RA
Local Access Control
Policy DB
PDP
STS
@neuFUSE
@neuRISK
@neuENDO
@neuLINK
Service Discovery
Data Transfer Service
Messaging
National Health - PKI

Service Providers

Local Access Control
Policy DB
PDP
STS
Compute Resources
Storage
Logging/Auditing
Data Access Service
@neuCOMPUTE Services
Data Transfer Service
Messaging
Data Staging Service
Logging/Auditing
@neuINFO
@neurIST RA

Research Institutes

Text Mining DB
Data Mining DB
Logging/Auditing
Data Access Service
Local Access Control
Policy DB
PDP
STS
Data Staging Service
Messaging
@neuLINK
@neuLINK application
Public DBs
@neurIST RA

Device Manufacturers & Consultants

@neurIST RA
Stent DB
Logging/Auditing
Data Access Service
Local Access Control
Policy DB
PDP
STS
Data Staging Service
Messaging
@neuENDO
@neuENDO Application

RA – Registration Authority DB - Database
🟦 @neurIST Communication Middleware
🟩 Security/Access Control Infrastructure
🟪 @neurIST Application Suite

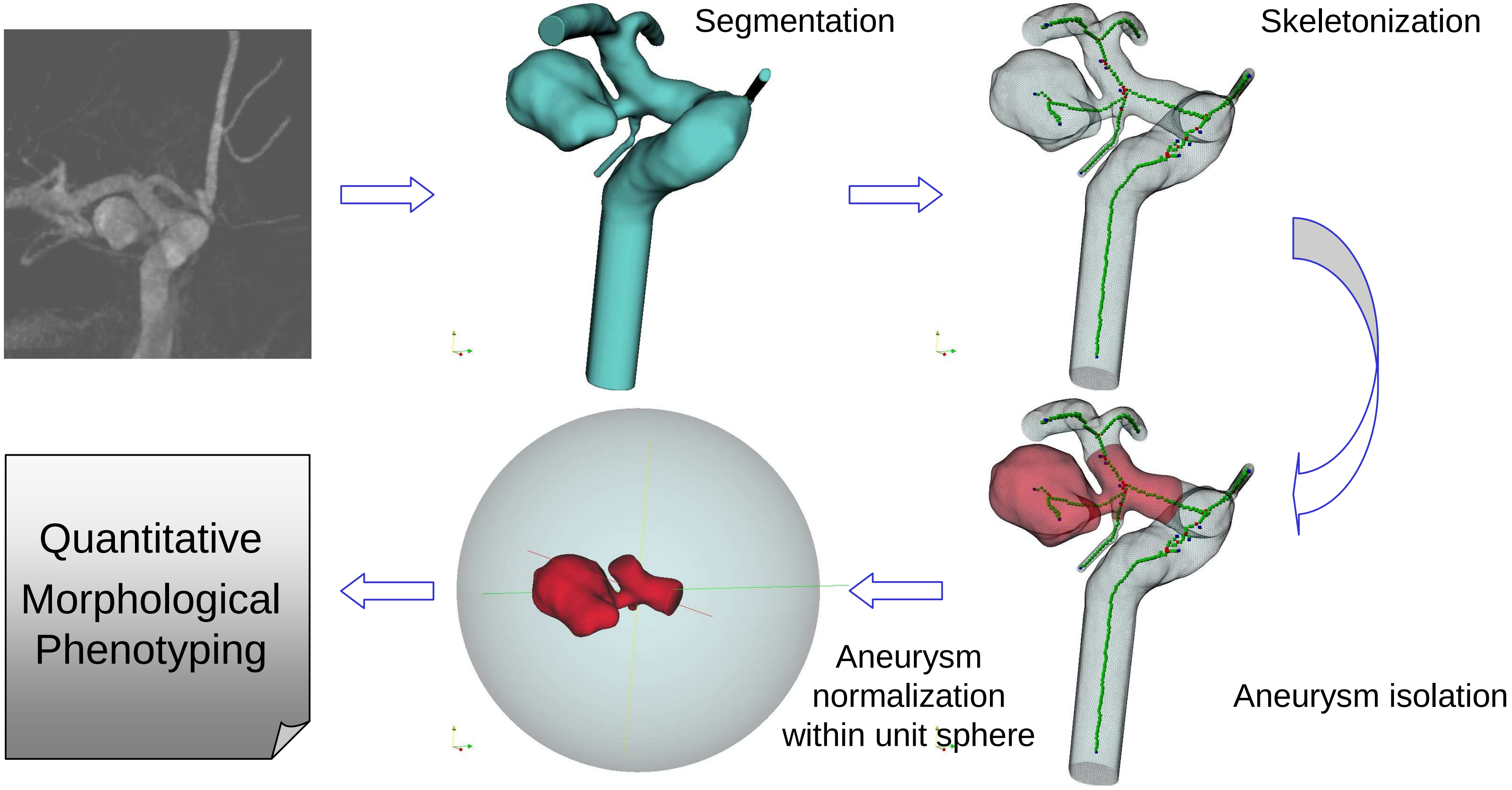
Clinical issue

Since aneurysm shape is likely to be dependent on the balance between haemodynamic forces and the aneurysmal surrounding environment, an appropriate morphological 3-D characterization is likely to provide a practical surrogate to quickly evaluate the risk of rupture.

Modeling this morphological characterization of all individuals in a large study population.

Compare morphological characterizations

Morphological phenotyping - global shape descriptor



Segmentation

Skeletonization

Aneurysm isolation

Aneurysm normalization within unit sphere

Quantitative Morphological Phenotyping

Calculation of similarity

- Segmentation of the intracranial aneurysm
- Use of 3D Zernike moments
  - compressed representation of shape
  - yields a numerical vector
- Additionally add nominal information, e.g. location and side
- Distance based on Gower-metric, allowing nominals and numerics

Problem

Use of ‘monolithic mining scripts’ leads to difficulties with:

- maintenance
- re-use of mining strategies
- user presentation of data flows
- use by non R users
- presentation of results to non R users

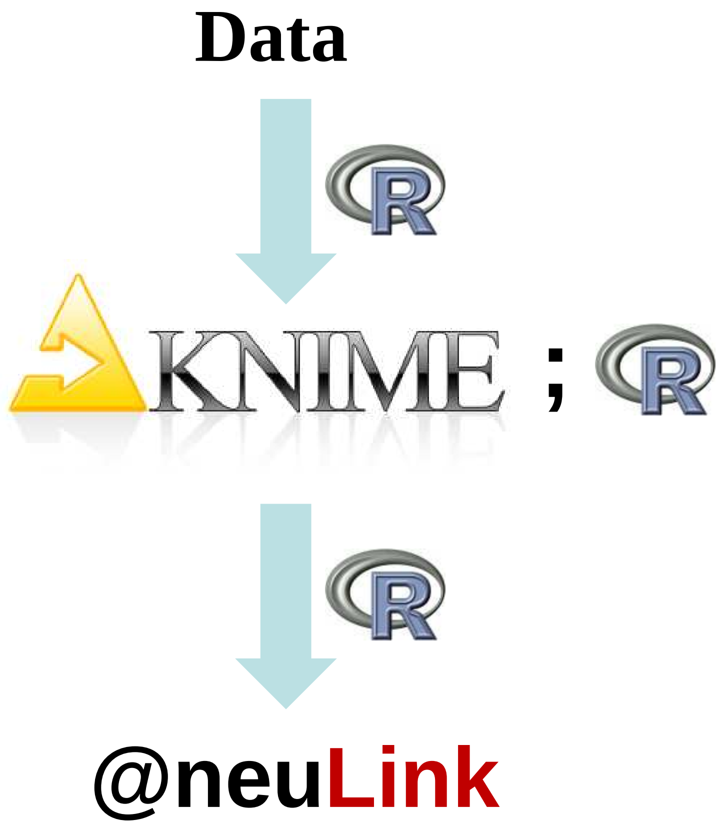
Roadmap to a solution

Considering that in the GRID context workflows like Taverna (http://taverna.sourceforge.net/) have shown great success (HealthGrid 2009, Medline publications);

Make use of a workflow engine:

- KNIME, exploration platform for multi-modal data:
  - R scripts
  - Weka
  - Java/Python
  - …
  - …

Overall workflow



Data

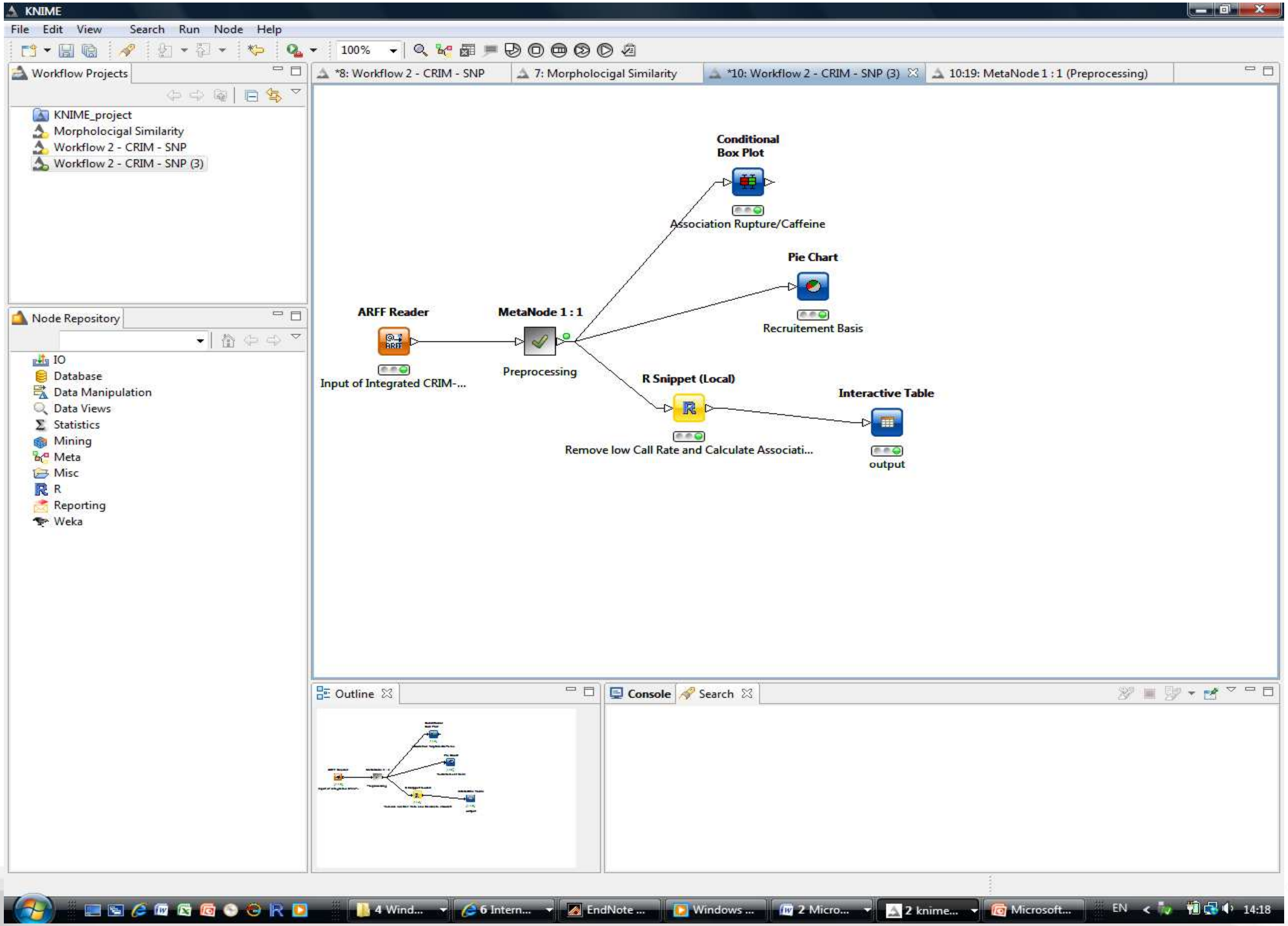
KNIME ; R

@neuLink

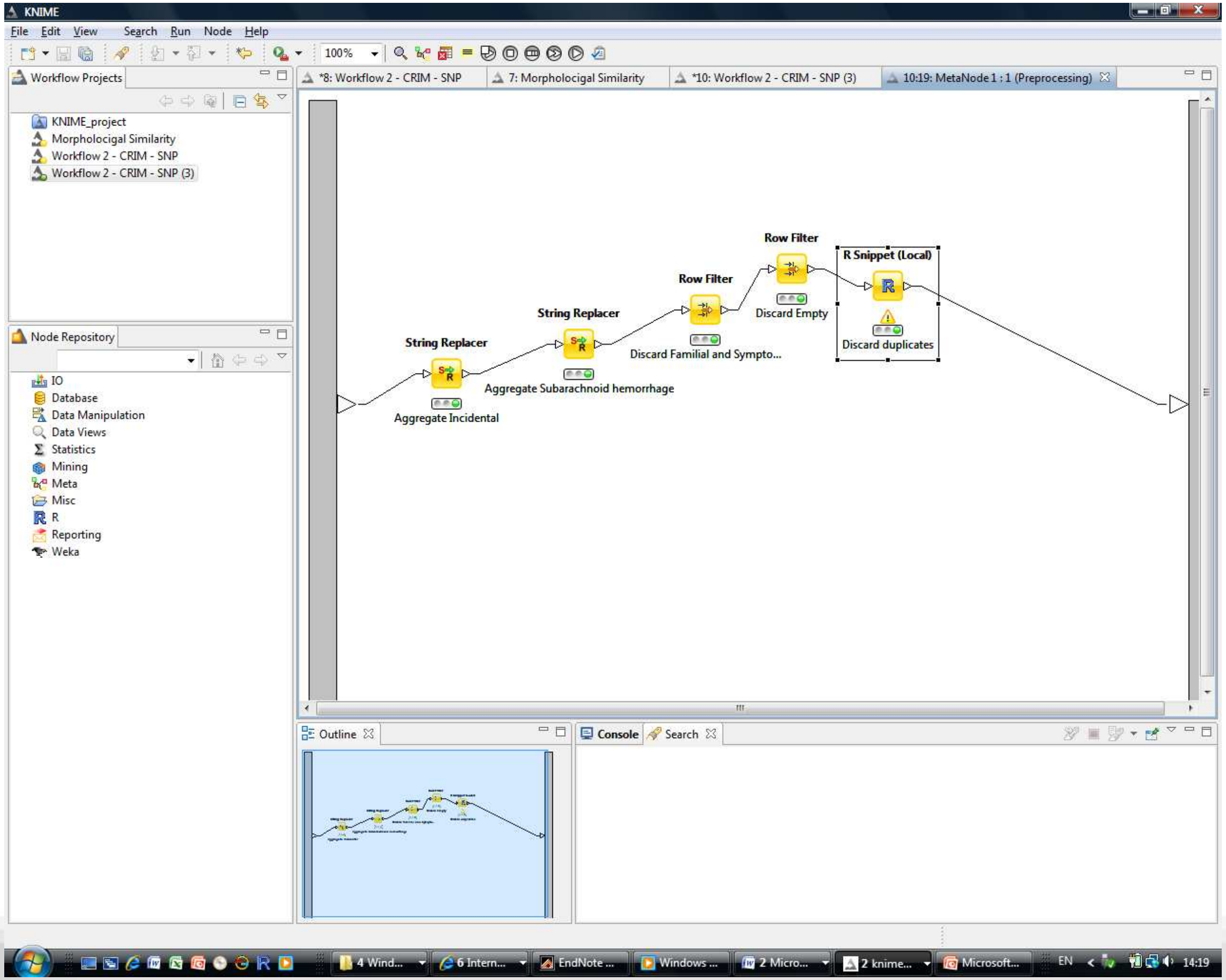
KNIME (http://www.knime.org)

- Konstanz Information Miner
- Workflow based Knowledge Discovery Tool
- Easy to use and Open Source
- Based on Eclipse
- Supports sub-workflows and evaluation workflows
- Includes ao R, Weka, and Java/Python
- Includes simple Graphics capabilities
- Workflows can be saved and run as a Batch

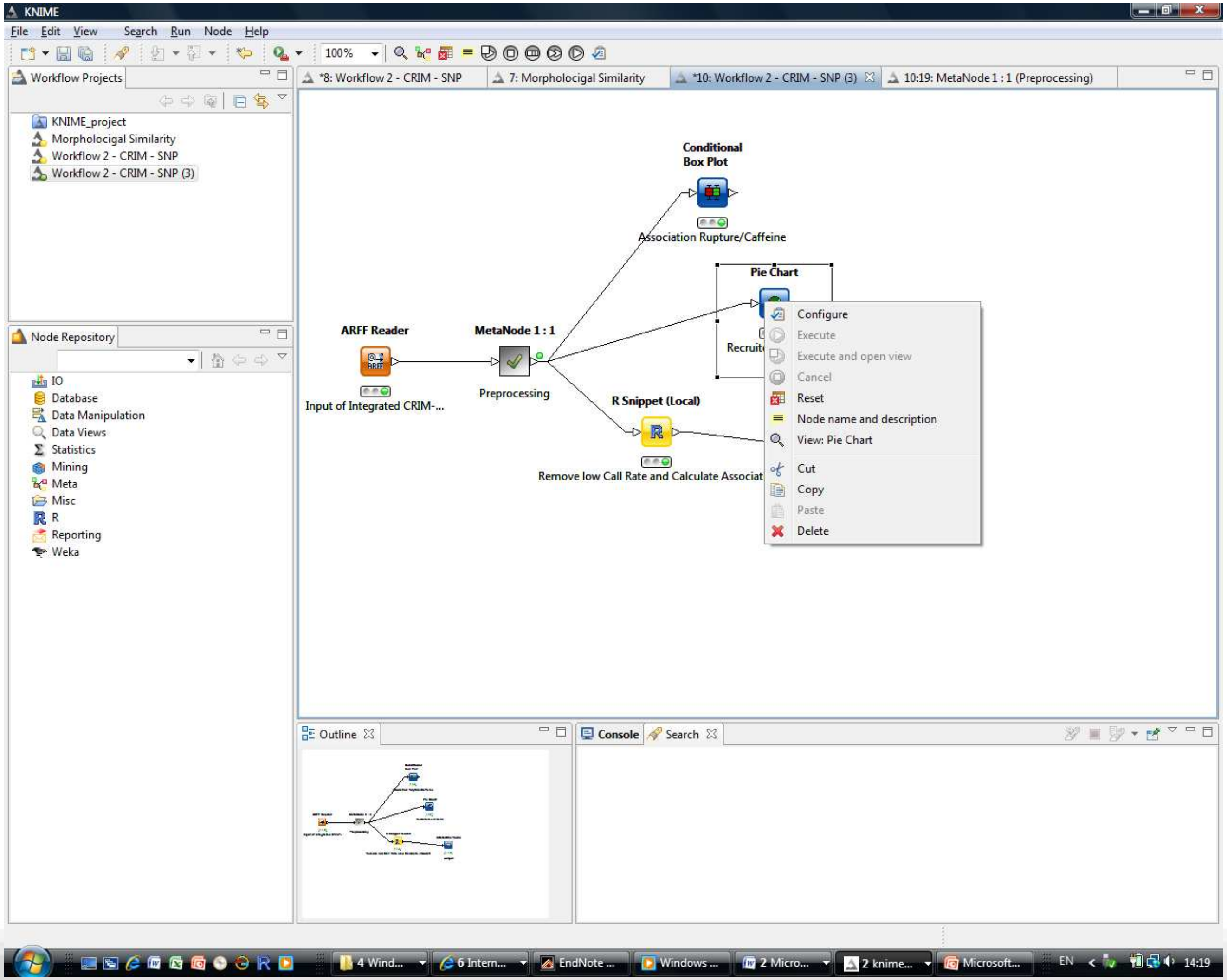
KNIME – Workflow overview



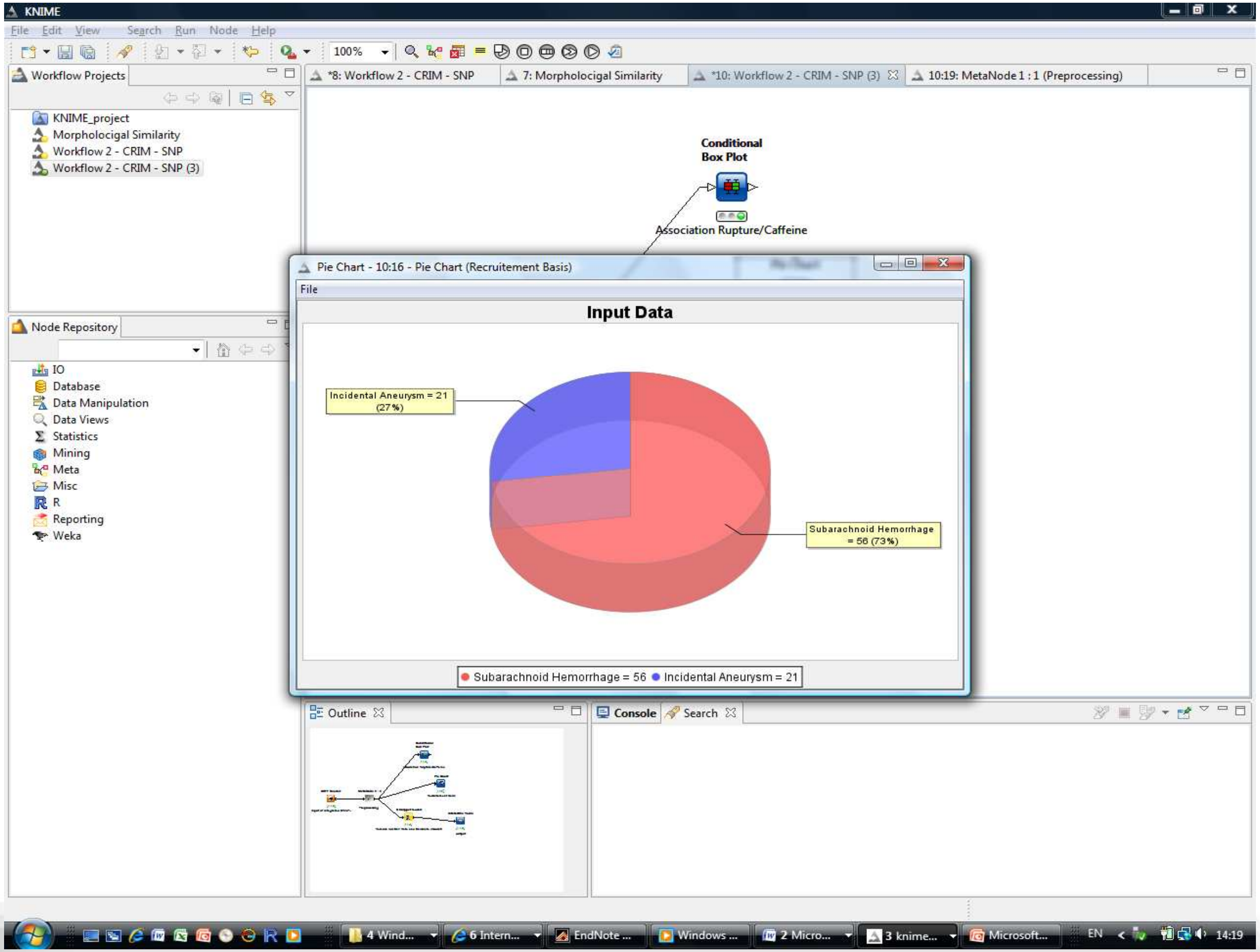
Meta node


KNIME
File Edit View Search Run Node Help
100%
Workflow Projects
KNIME_project
Morphological Similarity
Workflow 2 - CRIM - SNP
Workflow 2 - CRIM - SNP (3)
*8: Workflow 2 - CRIM - SNP
7: Morphological Similarity
*10: Workflow 2 - CRIM - SNP (3)
10:19: MetaNode 1 : 1 (Preprocessing)
String Replacer
Aggregate Incidental
String Replacer
Aggregate Subarachnoid hemorrhage
Row Filter
Discard Familial and Sympto...
Row Filter
Discard Empty
R Snippet (Local)
Discard duplicates
Node Repository
IO
Database
Data Manipulation
Data Views
Statistics
Mining
Meta
Misc
R
Reporting
Weka
Outline
Console
Search
4 Wind...
6 Intern...
EndNote ...
Windows ...
2 Micro...
2 knime...
Microsoft...
EN
14:19

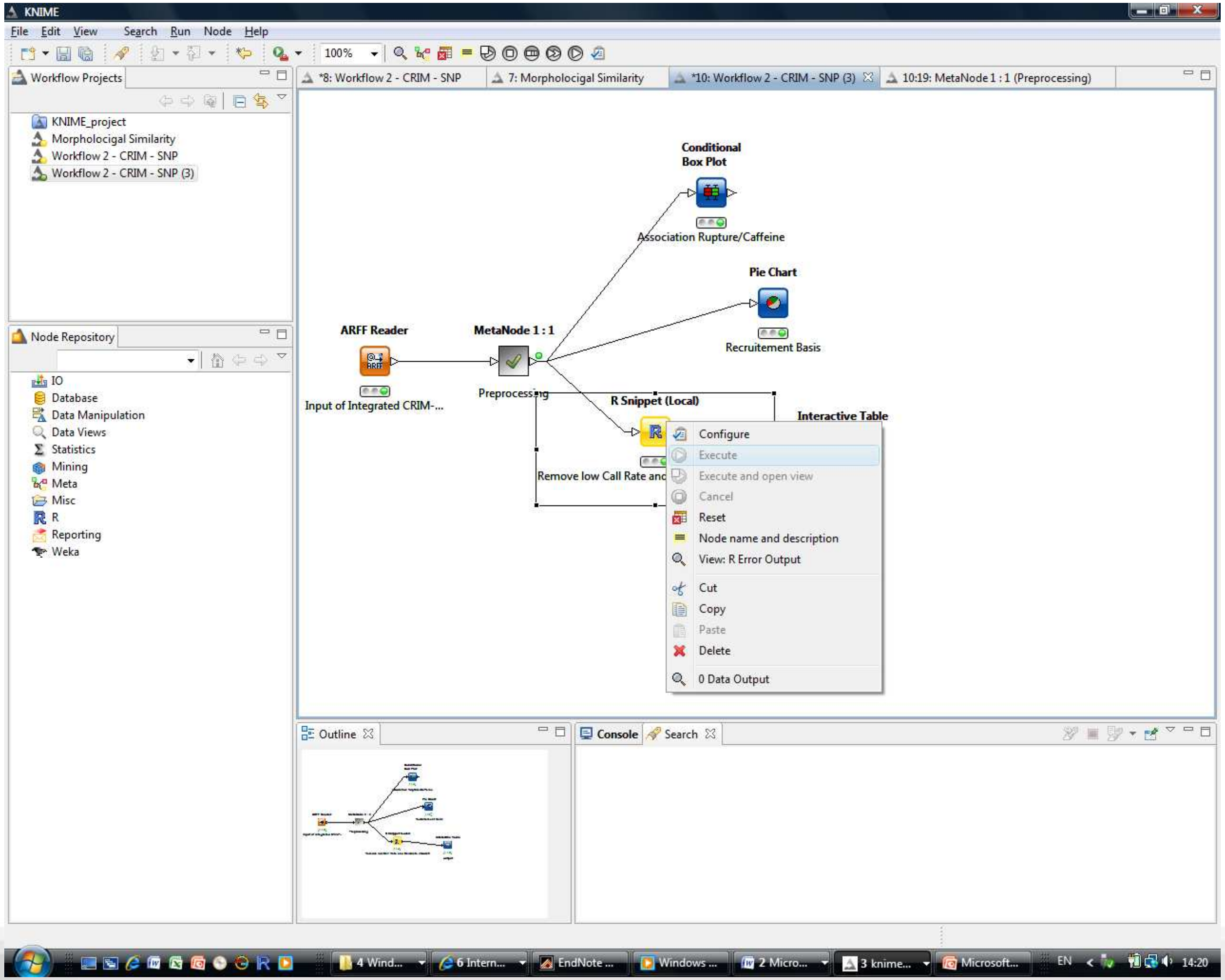
Click node to view Pie chart



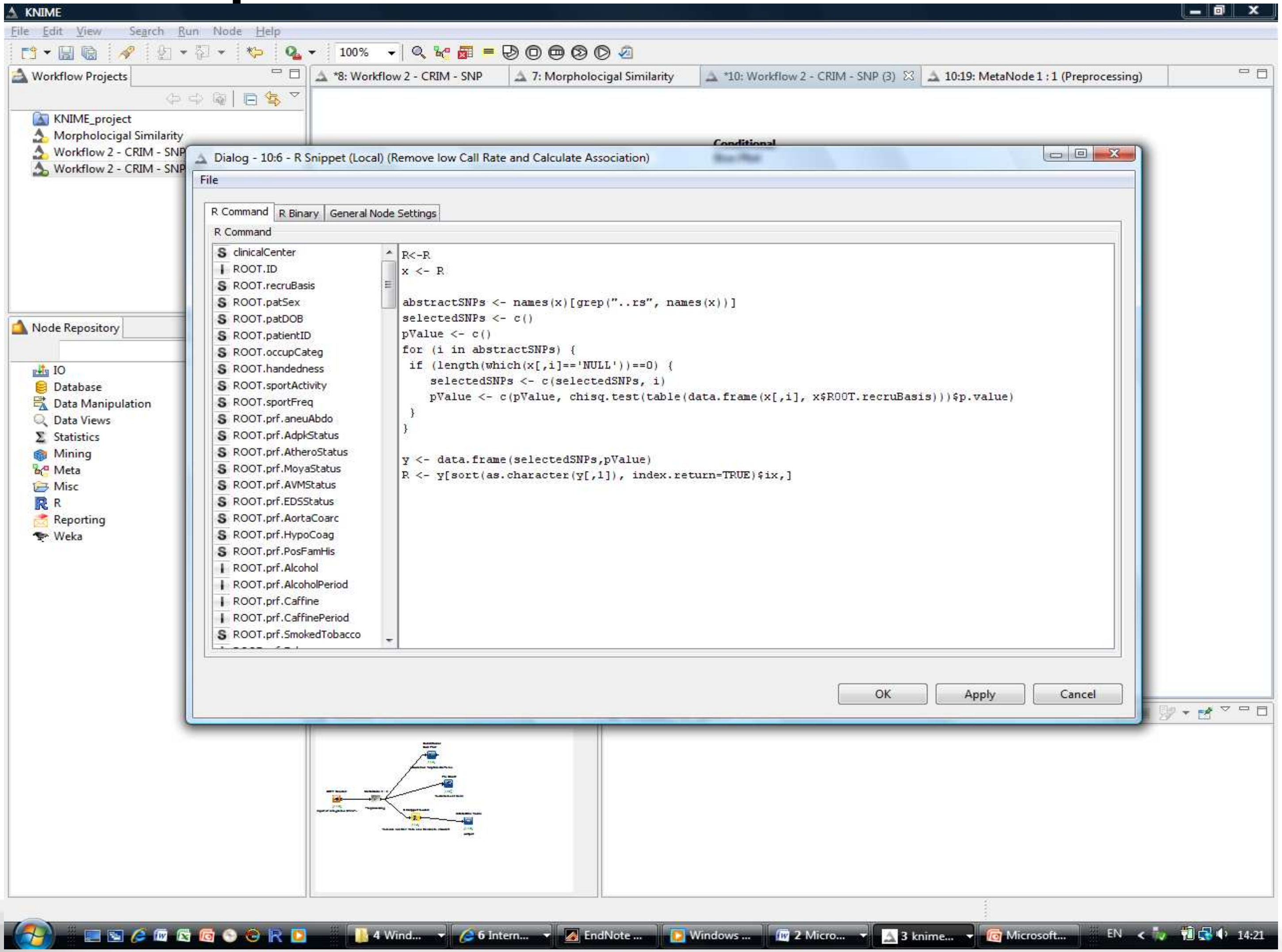
Pie chart


KNIME
File Edit View Search Run Node Help
100%
Workflow Projects
KNIME_project
Morpholocigal Similarity
Workflow 2 - CRIM - SNP
Workflow 2 - CRIM - SNP (3)
*8: Workflow 2 - CRIM - SNP
7: Morpholocigal Similarity
*10: Workflow 2 - CRIM - SNP (3)
10:19: MetaNode 1 : 1 (Preprocessing)
Conditional Box Plot
Association Rupture/Caffeine
Node Repository
IO
Database
Data Manipulation
Data Views
Statistics
Mining
Meta
Misc
R
Reporting
Weka
Pie Chart - 10:16 - Pie Chart (Recruitement Basis)
File
Input Data
Incidental Aneurysm = 21 (27%)
Subarachnoid Hemorrhage = 56 (73%)
Subarachnoid Hemorrhage = 56
Incidental Aneurysm = 21
Outline
Console
Search
4 Wind...
6 Intern...
EndNote ...
Windows ...
2 Micro...
3 knime...
Microsoft...
EN
14:19

Click R node to view script



R-script

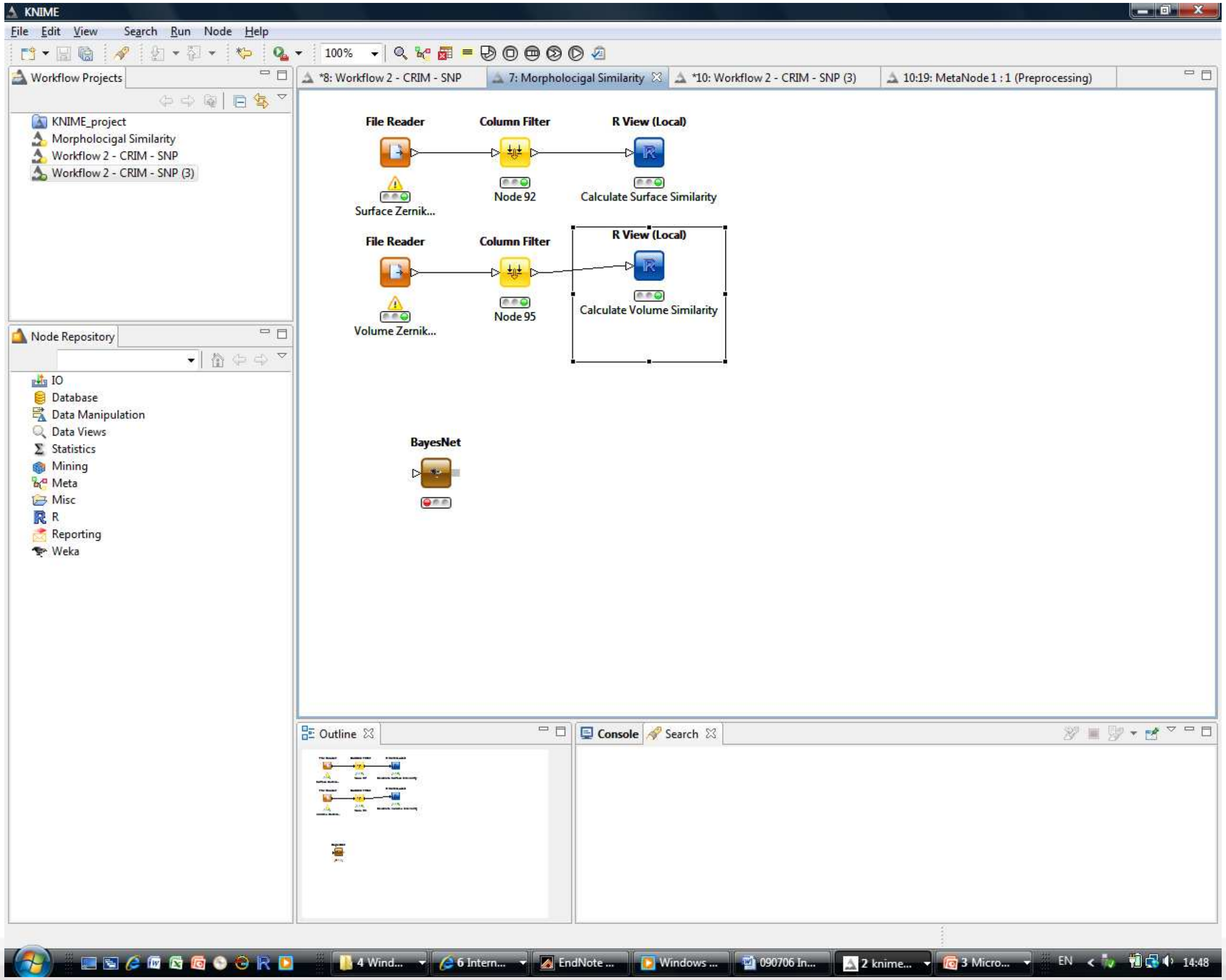


```
R<-R
x <- R

abstractSNPs <- names(x)[grep("..rs", names(x))]
selectedSNPs <- c()
pValue <- c()
for (i in abstractSNPs) {
 if (length(which(x[,i]=='NULL'))==0) {
    selectedSNPs <- c(selectedSNPs, i)
    pValue <- c(pValue, chisq.test(table(data.frame(x[,i], x$ROOT.recruBasis)))$p.value)
 }
}

y <- data.frame(selectedSNPs,pValue)
R <- y[sort(as.character(y[,1]), index.return=TRUE)$ix,]
```

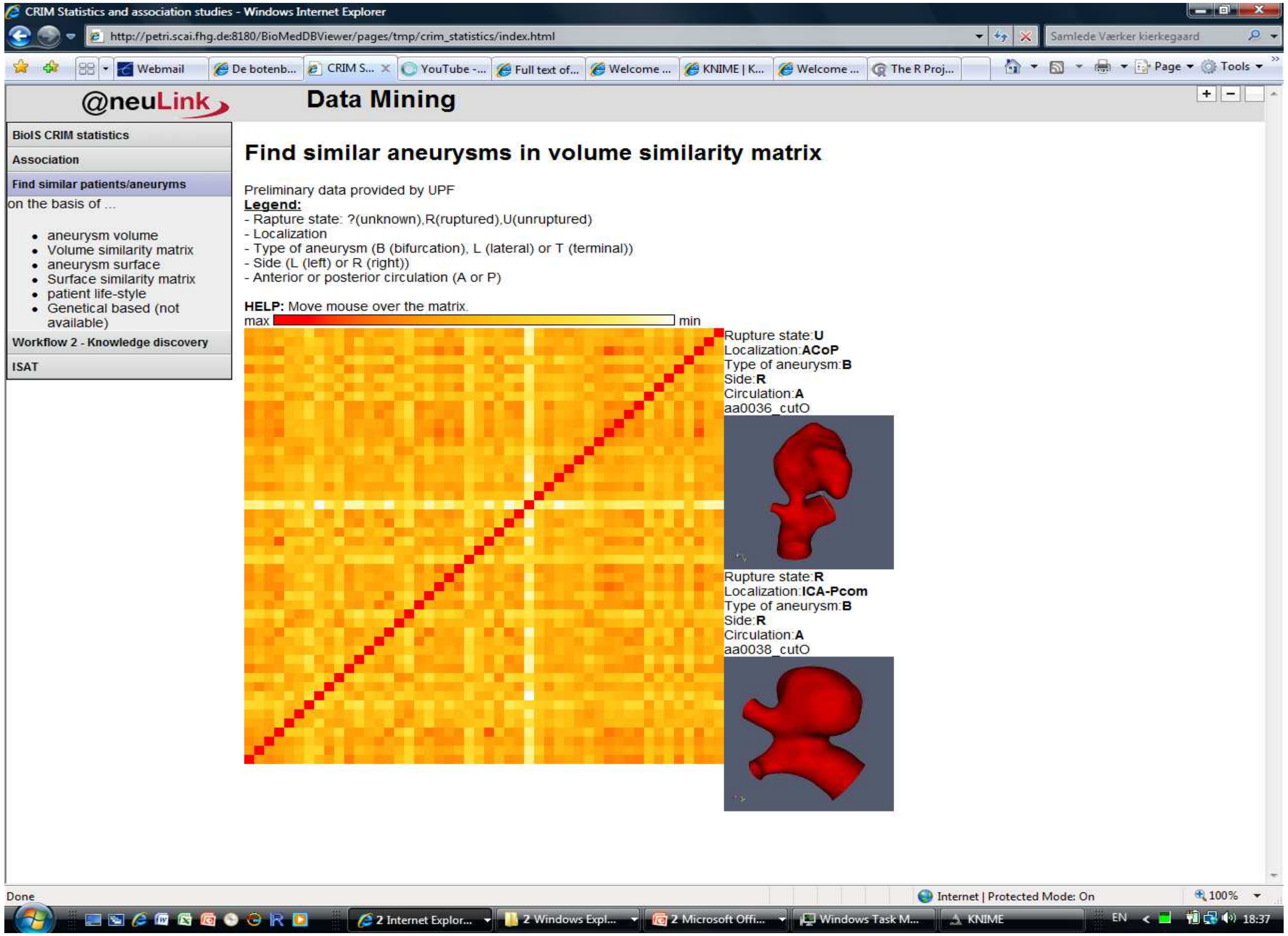
Volume based symmetry workflow



R script

```
for (i in 1:length(blo.dis)) {
   inter <-  as.character(blo.dis[sort(dis.matrix[i,], index.return=TRUE)$ix[1:10]])
   distan <- dis.matrix[i,sort(dis.matrix[i,], index.return=TRUE)$ix[1:10]]
   inter <- as.vector(rbind(inter, distan))
   blo.res <- rbind(blo.res, inter)  }
write.table(blo.res, paste("similarAneurysmsId",modeName,".txt", sep=""), quote=FALSE,
       row.names=FALSE, sep="\t", col.names=FALSE)
write.table(dis.matrix, paste("similarityMatrix",modeName,".txt", sep=""), quote=FALSE,
       row.names=FALSE, sep="\t", col.names=FALSE)
png("similarAneurysmsClusterSurface.png")
 image(1:nrow(dis.matrix), 1:nrow(dis.matrix), dis.matrix, main=paste(modeName," based
       Similarity", sep=""))
dev.off()
x.res <- biclust(dis.matrix, method=BCCC(), number=2)
drawHeatmap(dis.matrix,x.res,1)
image(1:nrow(dis.matrix), 1:nrow(dis.matrix), dis.matrix, main=paste(modeName," based
       Similarity", sep=""))
setwd(savedWd)
```

Outcome of similarity to @neuLink



on the basis of ...

- aneurysm volume
- Volume similarity matrix
- aneurysm surface
- Surface similarity matrix
- patient life-style
- Genetical based (not available)

Find similar aneurysms in volume similarity matrix

Preliminary data provided by UPF

Legend:
- Rapture state: ?(unknown),R(ruptured),U(unruptured)
- Localization
- Type of aneurysm (B (bifurcation), L (lateral) or T (terminal))
- Side (L (left) or R (right))
- Anterior or posterior circulation (A or P)

HELP: Move mouse over the matrix.

Rupture state:**U**
Localization:**ACoP**
Type of aneurysm:**B**
Side:**R**
Circulation:**A**
aa0036_cutO

Rupture state:**R**
Localization:**ICA-Pcom**
Type of aneurysm:**B**
Side:**R**
Circulation:**A**
aa0038_cutO

References

- Rajasekaran, H.; , Iacono, L. L.; Hasselmeyer, P.; Fingberg, J.; Summers, P.; Benkner, S.; Engelbrecht, G.; Arbona, A.; Chiarini, A.; Friedrich, C. M.; Hofmann-Apitius, M.; Kumpf, K.; Moore, B.; Bijlenga, P.; Iavindrasana, J.; Mueller, H.; Hose, R. D.; Dunlop, R., and Frangi, A. *@neurIST – Towards a System Architecture for Advanced Disease Management through Integration of Heterogeneous Data, Computing, and Complex Processing Services*. Proceedings of the 21st IEEE International Symposium on Computer-based Medical Systems, IEEE, 2008, 361-366.
- Friedrich, C. M.; Dach, H.; Gattermayer, T.; Engelbrecht, G.; Benkner, S., and Hofmann-Apitius, M. *@neuLink: A Service-oriented Application for Biomedical Knowledge Discovery*. Proceedings of the HealthGrid 2008, IOS Press, 2008, 165-172
- Berthold, M. (2009). KNIME (Konstanz Information Miner) webpage; last accessed 2009-02-26 http://www.knime.org/
- Witten, I. H., and Frank, E. *Data Mining*; second edition; Morgan Kaufmann Publishers, 2005
- Millán, R. D.; Dempere-Marco, L.; Pozo, J. M.; Cebral, J. R., and Frangi, A. F. *Morphological characterization of intracranial aneurysms using 3-D moment invariants*. IEEE Trans Med Imaging, 2007, 26, 1270-1282
- J.C. Gower, *A General Coefficient of Similarity and Some of ist Properties*. Biometrics, 27(4), 857-871, 1971.
- Hernández M., Frangi A..F. (2006). *Non-parametric Geodesic Active Regions: Method and evaluation for cerebral aneurysms segmentation in 3DRA and CTA*. Medical Image Analysis, in press.

aneurIST

Integrated biomedical informatics for the management of cerebral aneurysms

Workflows for Data Mining in integrated multi-modal data on Intracranial Aneurysms using KNIME

Roelof Risselada, Christoph M. Friedrich, Christian Ebeling, Roman Klinger, Anna Bauer-Mehren, Manuel Pastor, Maria Cruz Villa, Jose M. Pozo, Alejandro F. Frangi, and Martin Hofmann-Apitius

UseR! Conference, 7 July 2009, Rennes, France

The financial support of the European Commission is gratefully acknowledged. Material in this presentation reflects only the author's views and the Commission is not liable for any use that may be made of the information contained herein.