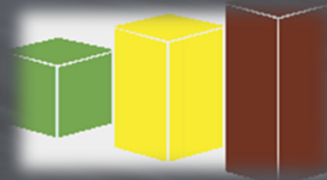




RESEARCH DATA ALLIANCE



W E L C O M E

## Interest Group **Big Data Analytics**

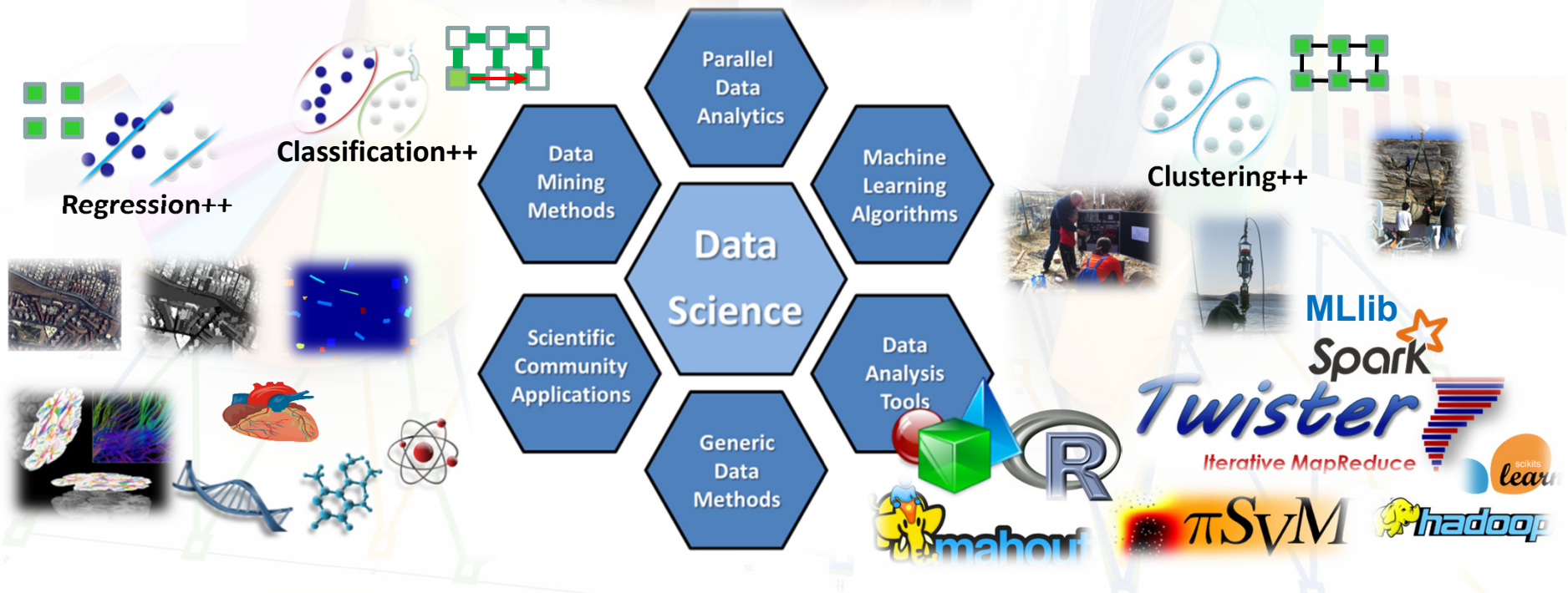
*Co-chairs Morris Riedel, Kuo Kwo-Sen, Peter Baumann*

research data sharing without barriers

**rd-alliance.org**

# Serial Algorithms for Large Volumes of Data Exist

## 'Big Data' Requires Parallel Algorithms and Open Availability







# Towards Systematic Data Analytics

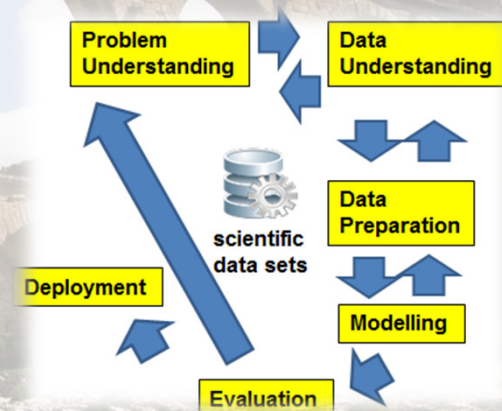
Guided by the Cross Industry Standard Process  
for Data Mining (CRISP-DM) Phases

*'Building a UCI Repository for Big Data Analytics'*



RESEARCH DATA ALLIANCE

Big Data Analytics IG  
Big Data Infrastructure WG



„Reference Data Analytics“  
for reusability & learning

CRISP-  
DM  
Report



Openly  
Shared  
Datasets



Running  
Analytics  
Code





## Results Example



RESEARCH DATA ALLIANCE

Big Data Analytics IG  
Big Data Infrastructure WG

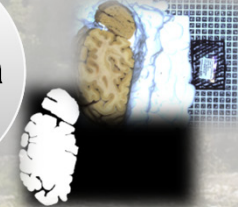
Future  
Grid

Twister  
Iterative MapReduce

$\pi$ SVM

Parallel  
Brain Data  
Analytics

scikits  
learn



„Reference Data Analytics“  
for reusability & learning

CRISP-  
DM  
Report



Openly  
Shared  
Datasets



Running  
Analytics  
Code



Sattelite Data(Quickbird)

Parallel  
Support Vector  
Machines (SVM)

HPC/MPI,  
Map-Reduce &  
GPGPUs



Classification  
Study of  
Land Cover  
Types

‘Best Practices’

Community-  
based practice

Classification++

G. Cavallaro and M. Riedel, ‘Smart Data Analytics  
Methods for Remote Sensing Applications’, IGARSS 2014