Giuseppe Longo\textsuperscript{1,4,5}

Ciro Donalek\textsuperscript{1,6}, Antonino Staiano\textsuperscript{1,4}
Roberto Tagliaferri\textsuperscript{2,5}, Angelo Ciaramella\textsuperscript{2,5}

1 - Dept. Of Physical Sciences, University of Napoli
2 – Dept. Of Mathematics, University of Salerno
3 - Caltech – Dept. Of Astronomy
4 - INAF –Napoli Unit
5 - INFN – Napoli Unit
6 – Caltech, Pasadena, USA

&

U. Becciani
INAF-Catania

C. Gheller
CINECA - Bologna
Discoveries

Data Mining
Visualization

Data Services

Existing Centers and Archives

DS6

we are Here - 2005
we were there - 2000

Djorgovski 2000, Munich
## What is data mining?

The capability to extract useful information from large data sets

<table>
<thead>
<tr>
<th>Guided by a priori knowledge</th>
<th>Based on statistical properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accurate</td>
<td>Less accurate</td>
</tr>
<tr>
<td>Reliable</td>
<td>More exciting</td>
</tr>
<tr>
<td>Socially acceptable</td>
<td>Socially questioned</td>
</tr>
</tbody>
</table>

- Feature selection
- Clustering
- Pattern and trend recognition
Feature selection
From an ocean of parameters, extract only those which are relevant to a given task

\[\textit{compression of data sets, computational speed}\]

Clustering:
the mathematical word for most of the tasks performed by the astronomers in interpreting data

Pattern and trend recognition:
All the other tasks done by the astronomers
IMPLEMENTATION OF AN AUTOMATIC CLASSIFIER (S/U) FOR MULTIBAND SURVEY DATA

1. Construction of the “a priori” knowledge data archive (first: existing data; after: integration with simulations)

2. Implementation of a feature selection supervised tool

3. Application to test data sets (Sloan, DPOSS, ESO archive data)

4. Generalization for usage on a distributed architecture

CDS (“a priori” knowledge)
Edinburgh, Portsmouth (SDSS data)
Napoli (DPOSS and A.I. know how)
Edinburgh, Leicester, Portsmouth, Trieste (GRID know-how)
Catania (Visualization tools)
CDS

labeling
Feature selection
Data Mining and Analysis

Clusters and classification

Results

data providers

a priori knowledge Providers

CDS