Photometric Redshifts with DAME

O. Laurino, R. D'Abrusco
M. Brescia, G. Longo
&
DAME Working Group

VO-Day ... in Tour
@INAF

The general astrophysical problem

Due to new instruments and new diagnostic tools, the information volume grows exponentially

➡ ***Most data will never be seen by humans!***

The need for data storage, network, database-related technologies, standards, etc.

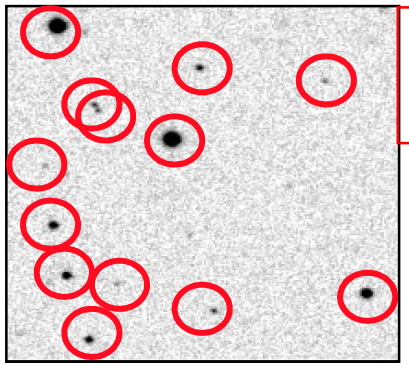
Information complexity is also increasing greatly

➡ ***Most knowledge hidden behind data complexity is lost***

Most (all) empirical relationships known so far depend on 3 parameters …. Simple universe or rather human bias?

➡ ***Most data (and data constructs) cannot be comprehended by humans directly!***

The need for data mining, KDD, data understanding technologies, hyperdimensional visualization, AI/Machine-assisted discovery



Detect sources and measure their attributes (brightness, position, shapes, etc.)

p={isophotal, petrosian, aperture magnitudes concentration indexes, shape parameters, etc.}

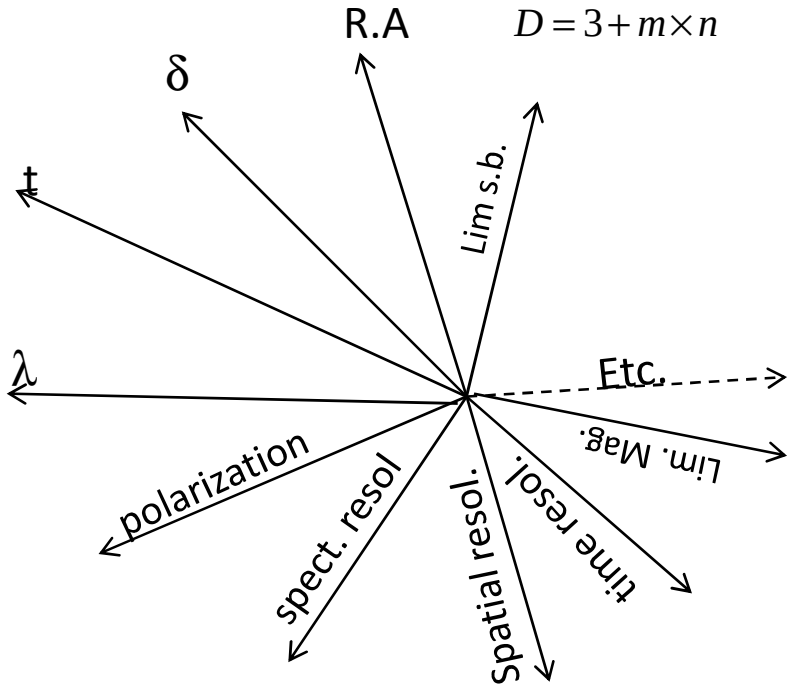
$$p^1 = \{RA^1, \delta^1, t, \{\lambda_1, \Delta\lambda_1, f_1^{1,1}, \Delta f_1^{1,1}, \ldots, f_1^{1,m}, \Delta f_1^{1,m}\}, \ldots, \{\lambda_n, \Delta\lambda_n, f_n^{1,1}, \Delta f_n^{1,1}, \ldots, f_n^{1,m}, \Delta f_n^{1,m}\}\}$$

$$p^2 = \{RA^2, \delta^2, t, \{\lambda_1, \Delta\lambda_1, f_1^{2,1}, \Delta f_1^{2,1}, \ldots, f_1^{2,m}, \Delta f_1^{2,m}\}, \ldots, \{\lambda_n, \Delta\lambda_n, f_n^{2,1}, \Delta f_n^{2,1}, \ldots, f_n^{2,m}, \Delta f_n^{2,m}\}\}$$

.....

$$p^N = \{RA^N, \delta^N, t, \{\lambda_1, \Delta\lambda_1, f_1^{N,1}, \Delta f_1^{N,1}, \ldots, f_1^{N,m}, \Delta f_1^{N,m}\}, \ldots\}$$

$$D = 3 + m \times n$$



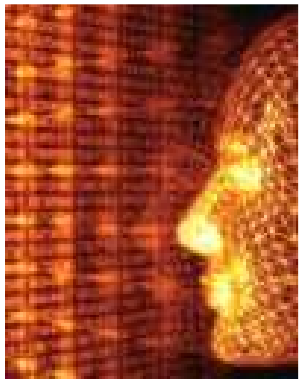
PARAMETER SPACE

From the Data Mining point of view, a**ny observed (simulated) datum *p* defines a point (region) in a subset of R^N.**

$$p \in \Re^N \qquad N >> 100$$

From data to knowledge: KDD

Knowledge Discovery in Databases



Data Gathering (e.g., from sensor networks, telescopes…)

→ Data Farming:

Storage/Archiving
Indexing, Searchability
Data Fusion, Interoperability, ontologies, etc.

Database technologies

→ Data Mining (or Knowledge Discovery in Databases):

Pattern or correlation search
Clustering analysis, automated classification
Outlier / anomaly searches
Hyperdimensional visualization

Key mathematical issues

→ Data understanding

Computer aided understanding
KDD
Etc.

Ongoing research

→ New Knowledge

Data Mining in the VO

• A new Interest group on Knowledge Discovery in Massive Data Sets was born inside the IVOA.
• The explosion of Data available (the “Data tsunami”) in the VO can be effectively dealth with using Data Mining, especially when the time variable comes into play.

But...The VO currently lacks general purpose tools for “server side” **massive data sets manipulation.**

Dame in the VO Framework

• To provide the VO with an extensible, integrated environment for **Data Mining and Exploration**;

• Support of the VO standards and formats, especially for application interop (SAMP);

• To abstract the application deployment and execution, so to provide the VO with an “opaque” general purpose computing platform taking advantage of the modern technologies (e.g. Grid, Cloud, etc...).

What is DAME

DAME is a joint effort between University Federico II, INAF-OACN, and Caltech aimed at implementing (as web application) a scientific gateway for data analysis, exploration, mining and visualization tools, on top of virtualized distributed computing environment.

http://voneural.na.infn.it/

Technical and management info

Documents

Science cases

Newsletter

http://dame.na.infn.it/

Web application PROTOTYPE

Photometric redshifts?

Multicolour photometry maps physical parameters:

luminosity L
redshift z
spectral type T

$\xrightarrow{f}$ observed fluxes

If the relation can be inverted then:

u,g,r,i,z,H,J,K,... $\xrightarrow{f^{-1}}$ **z**, L, T

The function can be approximated by regression in the photometric space. The accuracy of the photometric redshifts depend on two aspects:

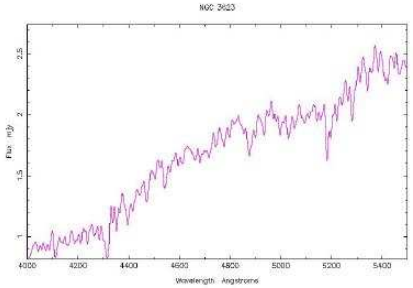
How the photometric filters cover the SED of the sources...

How absorption and emission features relate to the photometric filters...

A Science case: photometric redshifts (sample from SDSS galaxies)

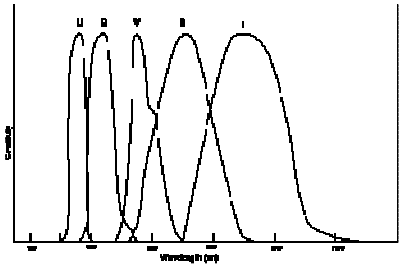
one of the main tools to investigate the spatial distribution of galaxies, i.e. to reconstruct the 3-D position of very large number of sources using only their photometric properties.

Spectral Energy Distribution convolved with band filters



Galaxy spectrum - F(λ)

X



Photometric system - $S_i(\lambda)$

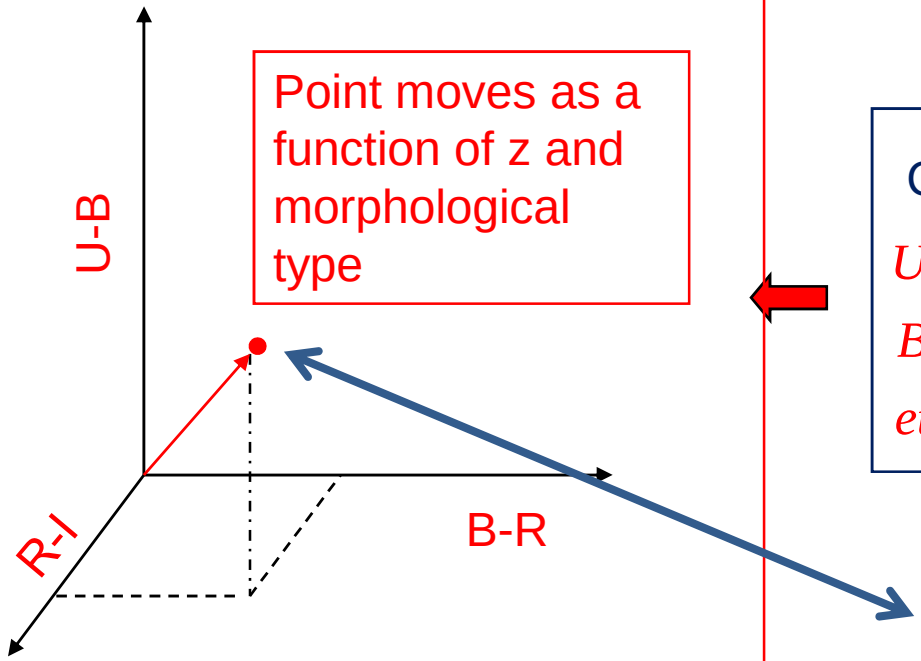
=

$$m_U = -2.5\log_{10}\frac{\int F(\lambda)S_U(\lambda)d\lambda}{\int S_U(\lambda)d\lambda} + c_u$$

$$m_B = -2.5\log_{10}\frac{\int F(\lambda)S_B(\lambda)d\lambda}{\int S_B(\lambda)d\lambda} + c_B$$

Etc…

Point moves as a function of z and morphological type



Color indexes

$U - B \equiv m_U - m_B$

$B - R \equiv m_B - m_R$

$etc.$

Photometric redshifts: the Data Mining approach

Photometric redshifts are treated as a regression problem (i.e. function approximation) in a multidimensional parameter space, hence a DM problem:

$\mathbf{X} \equiv \{x_1, x_2, x_3, \ldots x_N\}$ **input vectors**

$\mathbf{Y} \equiv \{x_1, x_2, x_3, \ldots x_M\}$ **target vectors** $M << N$

find $\hat{f}$: $\hat{\mathbf{Y}} = \hat{f}(\mathbf{X})$ **is a good approximation of** $\mathbf{Y}$

BoK = Base of Knowledge

BoK (from Vobs)
(set of templates)

→ Mapping function → Knowledge (phot-z's)

Observed Spectroscopic Redshifts

Synthetic colors from theoretical SEDs
Synthetic colors from observed SED's
.....

Knowledge always reflects the biases in the BoK.

Interpolative
Uneven coverage of parameter space

SED fitting
Unknown or oversimplified physics?
Unjustified assumptions!
Computational intensive!
Possibly accurate, but is it worth?
.....

DA ME

Photometric Redshifts with Neural Networks

• Spectroscopic observations are the most accurate method to determine redshifts, but time consuming;

• Photometric sources often outnumber Spectroscopic ones up to 3 orders of magnitude (it may depend on the BoK);

• If we build a reliable BoK with spectroscopic data we can reproduce the functional mapping between photometric parameters and redshift;

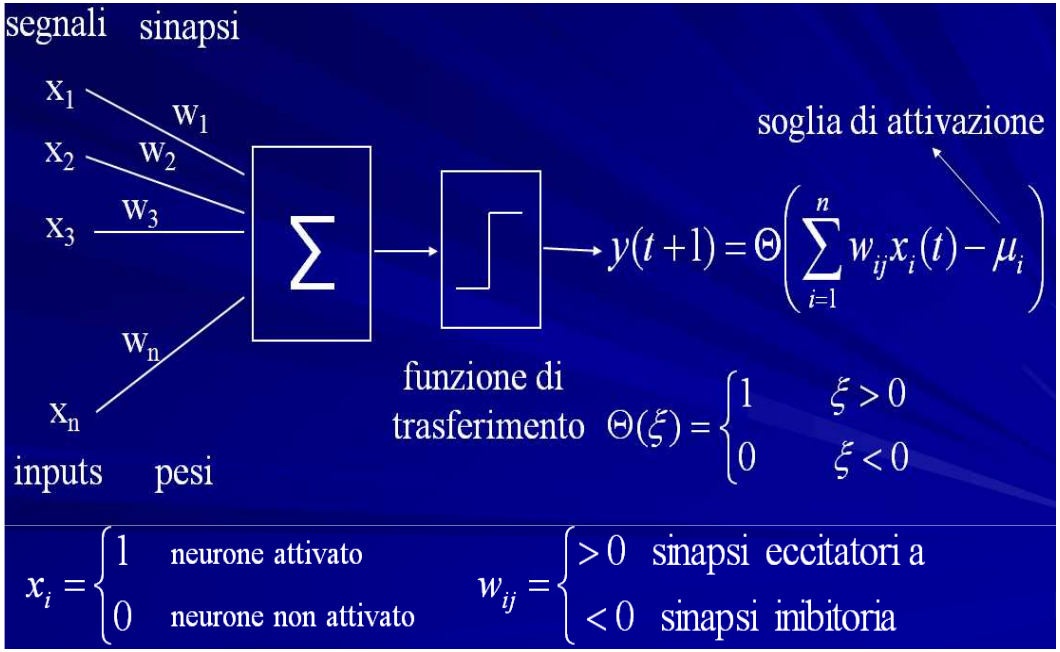
• Zphot accuracy is adequate for several astronomical applications;

Neural Networks advantages:

Fast;
Scale much better than any other method;
Learn by examples (BoK, in this case SDSS) and adapts easily to new data;
Do not require *a priori assumption on the Spectral Energy Distributions of sources;*
May be applied to all classes of extragalactic sources;

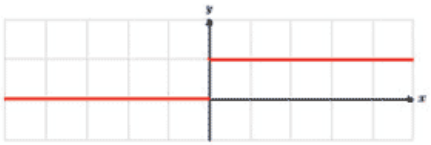
MLP – The Architecture

Mc-Culloch & Pitts artificial Neuron model



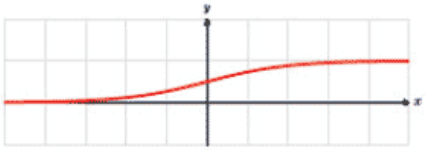
Neuron activation function

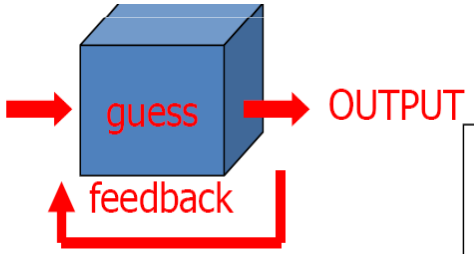
Gradino $\Phi(A) = \begin{cases} 1 & A > \theta \\ 0 & altrimenti \end{cases}$

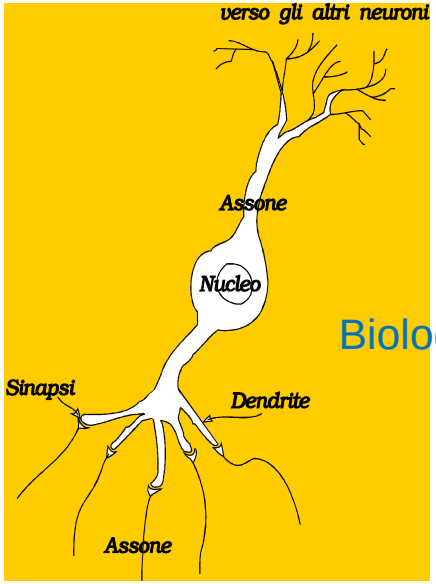


Lineare $\Phi(A) = kA$

Sigmoide $\Phi(A) = \frac{1}{1+e^{-kA}}$





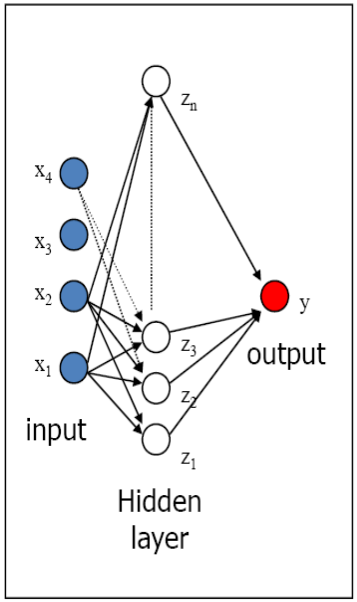


Biological Neuron

- input layer (n neurons)
- M hidden layer (1 or 2)
- Output layer (n' <n neurons)

Neurons are connected via activation functions

Different NN's given by different topologies, different activation functions, etc.

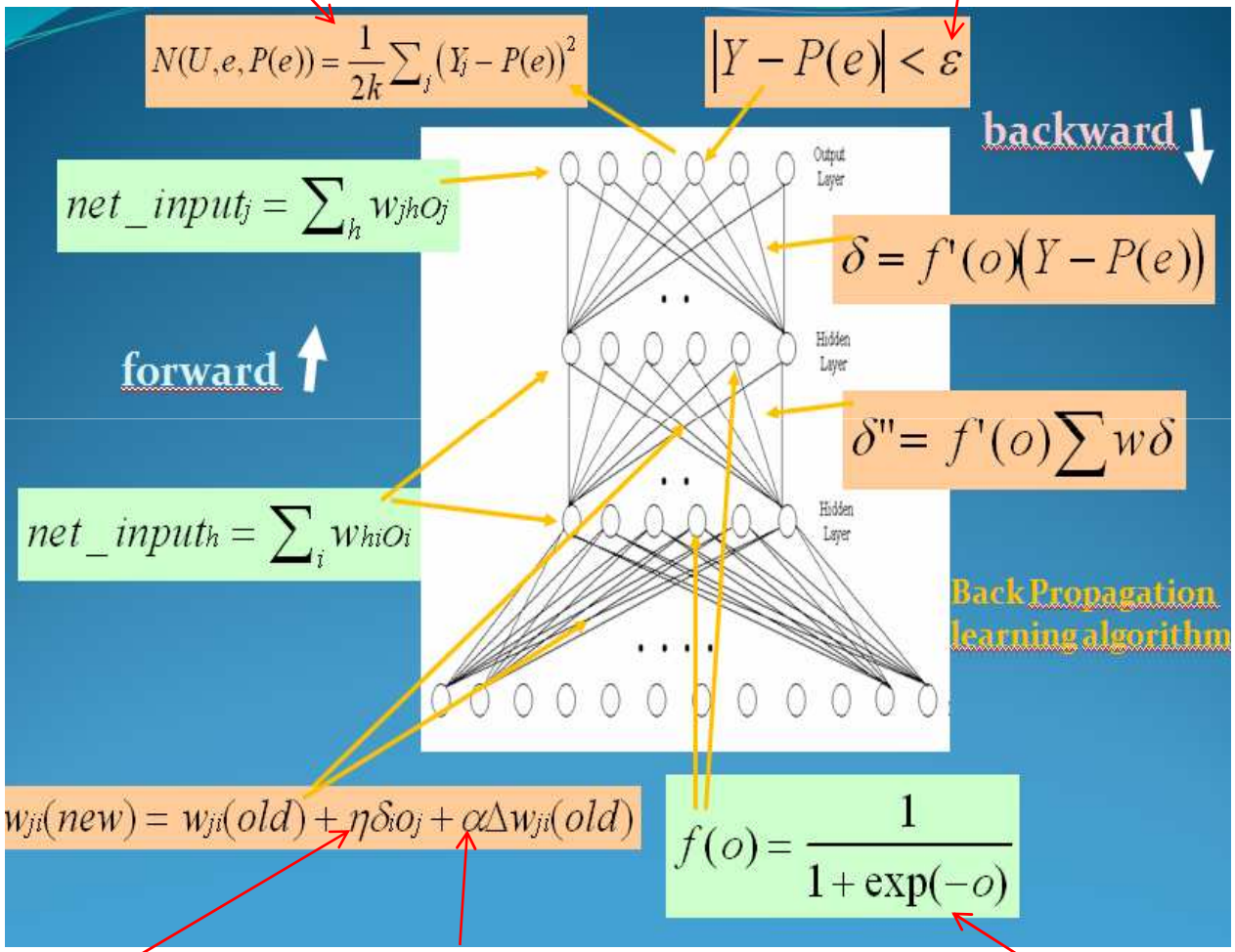


MLP – Back Propagation learning algorithm

Output error

Error tolerance threshold

FeedForward phase with input propagated through neuron activation function calculated at each layer



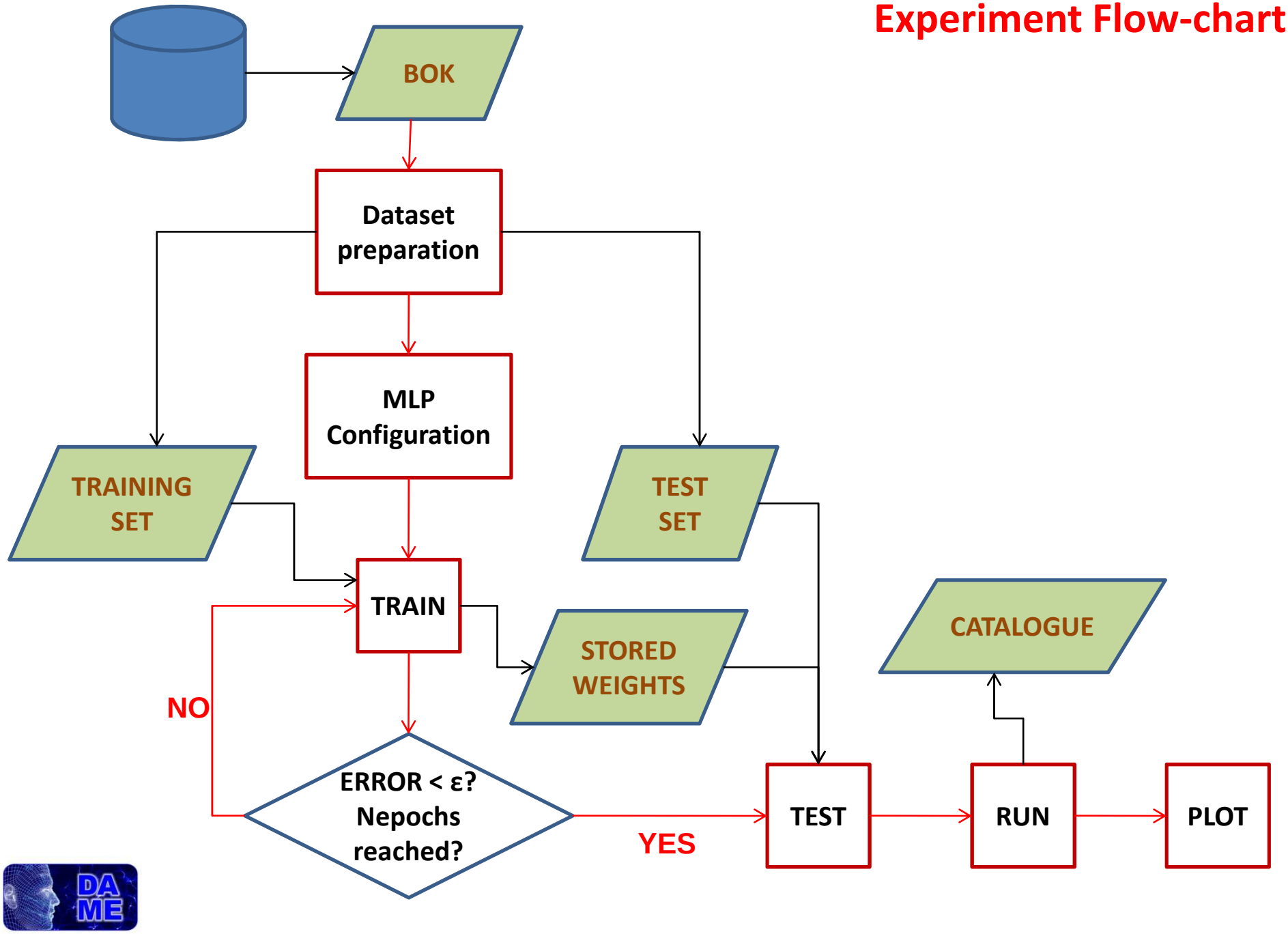
Backward phase with activation function gradient calculation and back propagation of learning error through layers, in order to update internal weights

Weight updating law

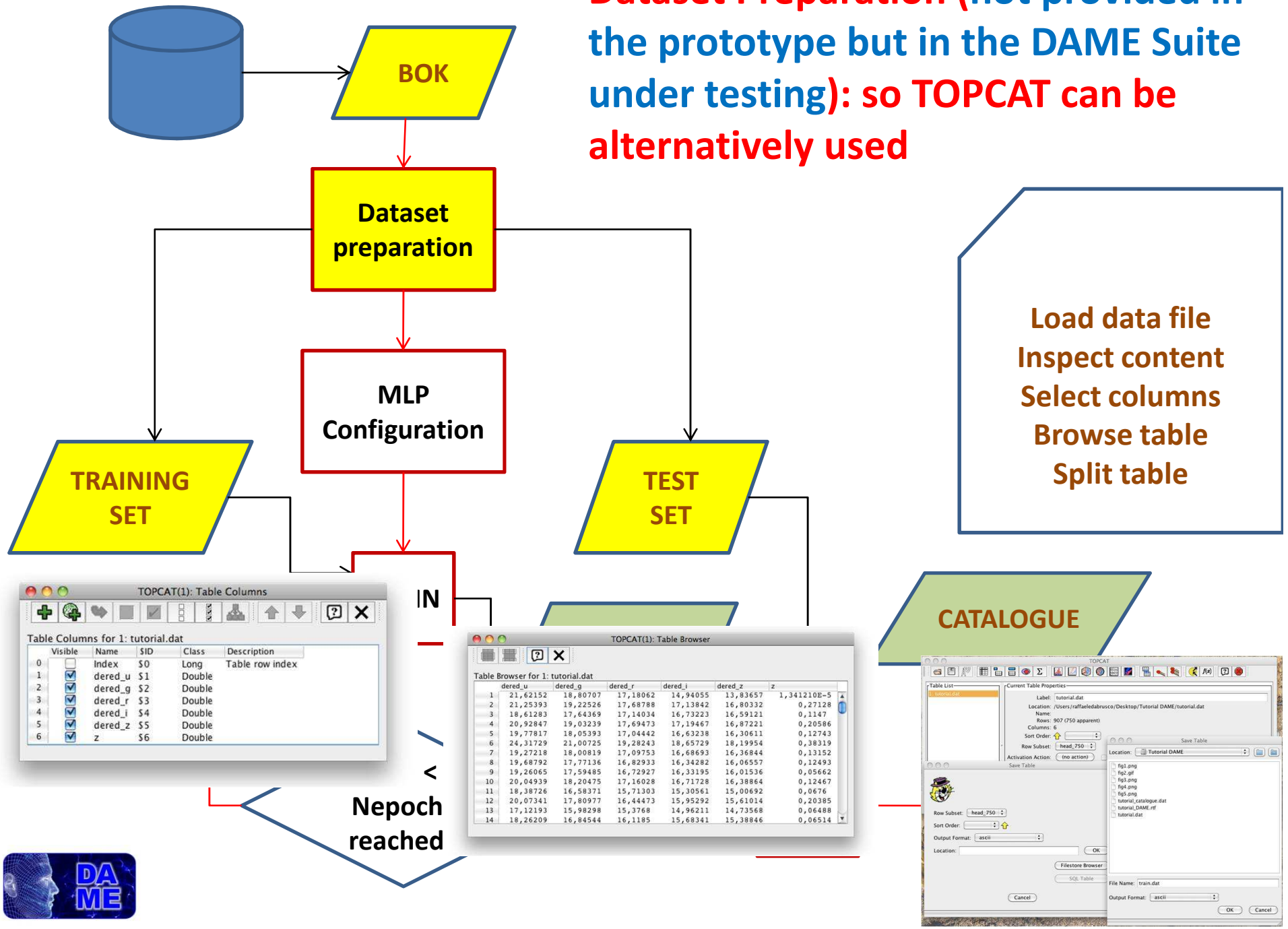
learning rate

Activation function

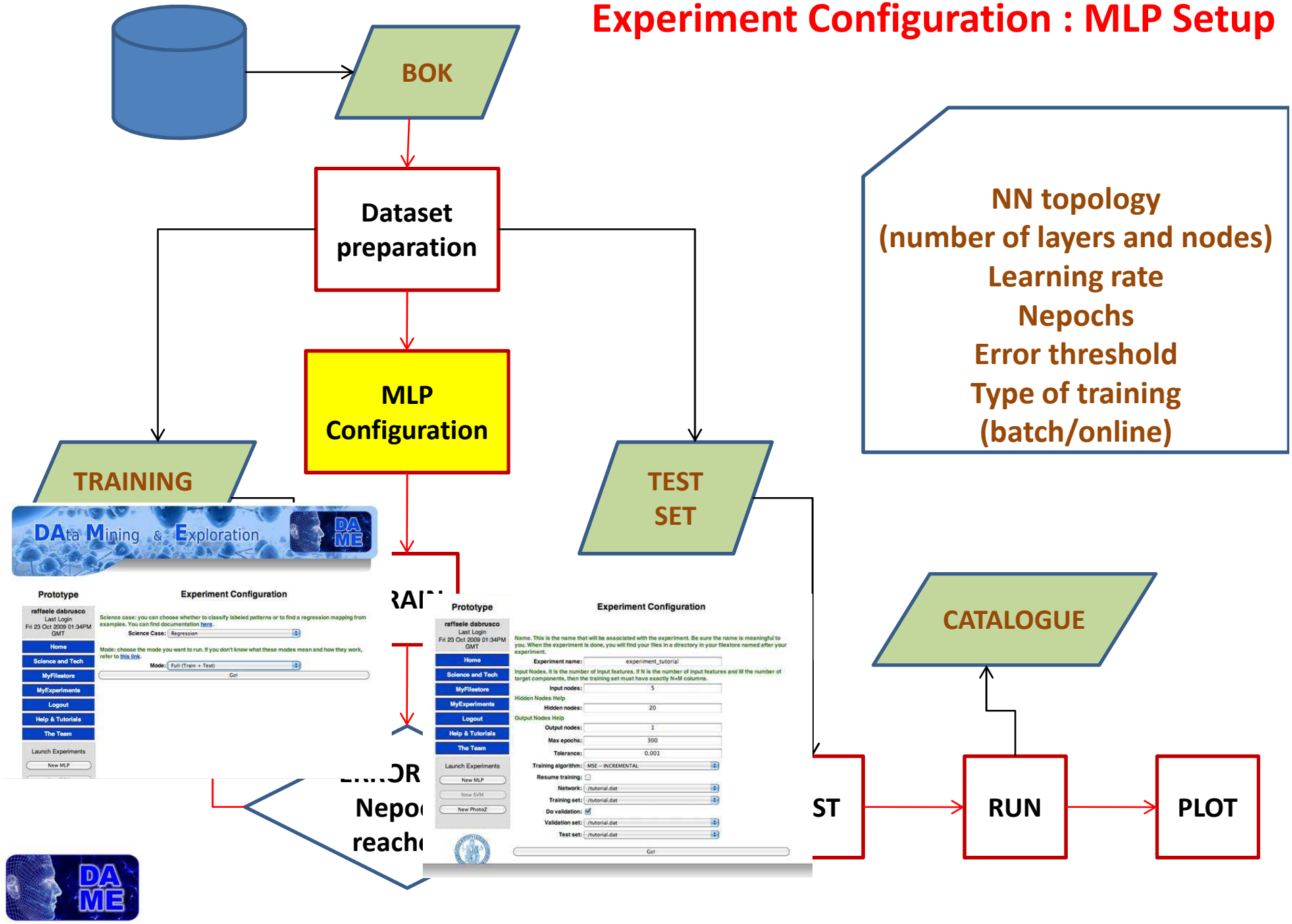
Experiment Flow-chart



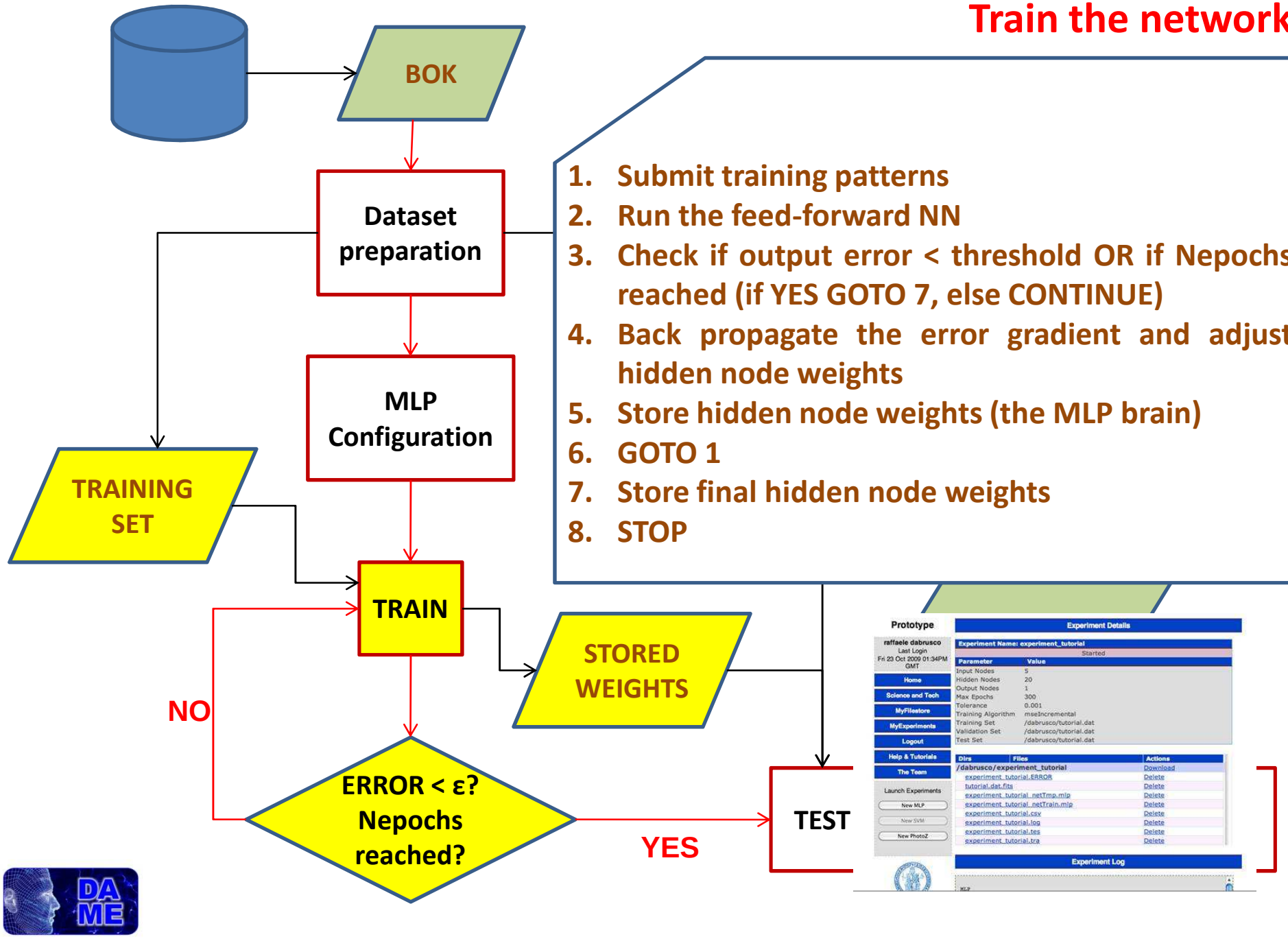
Dataset Preparation (not provided in the prototype but in the DAME Suite under testing): so TOPCAT can be alternatively used



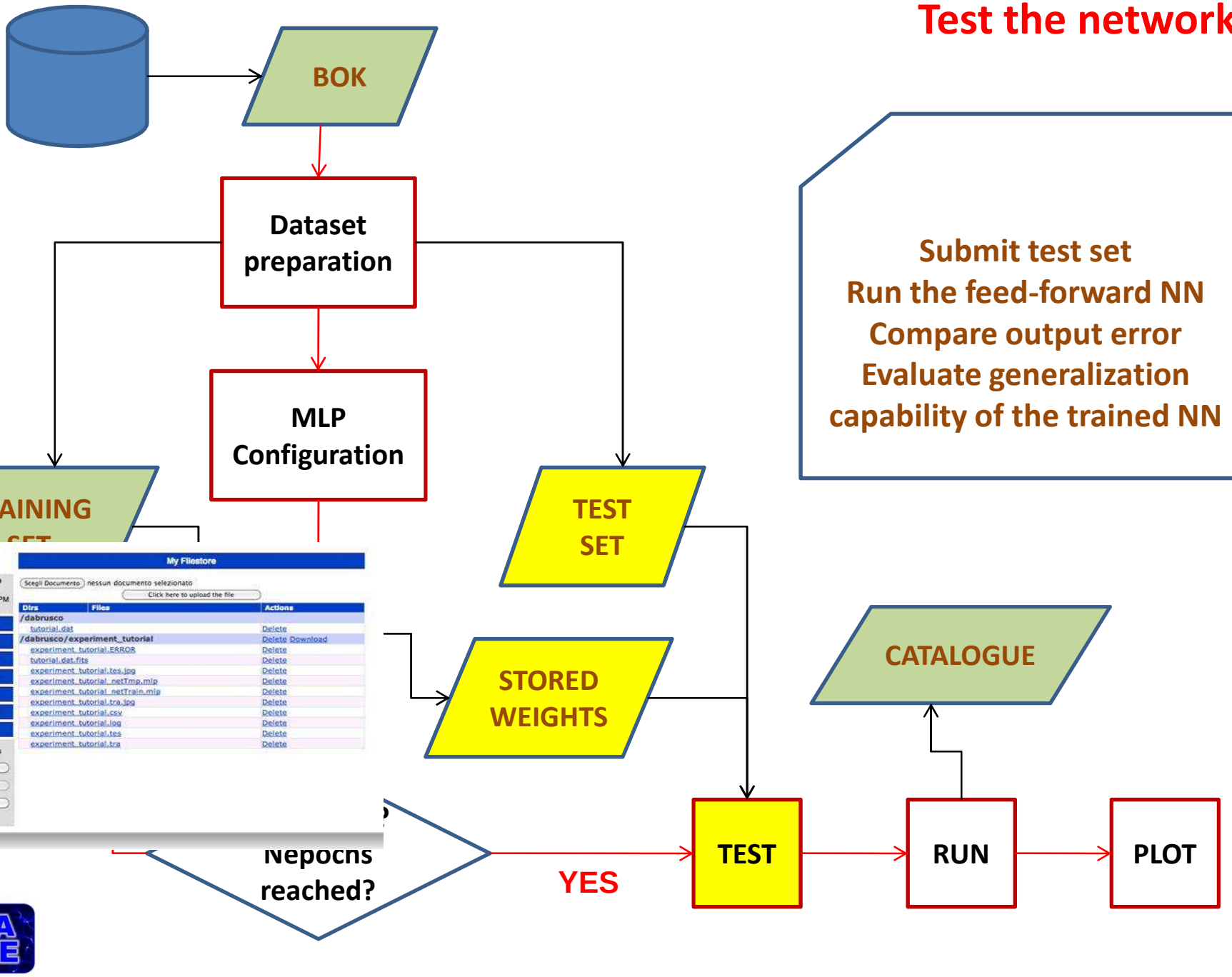
Experiment Configuration : MLP Setup



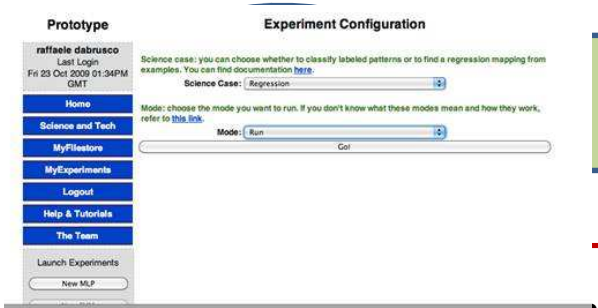
Train the network



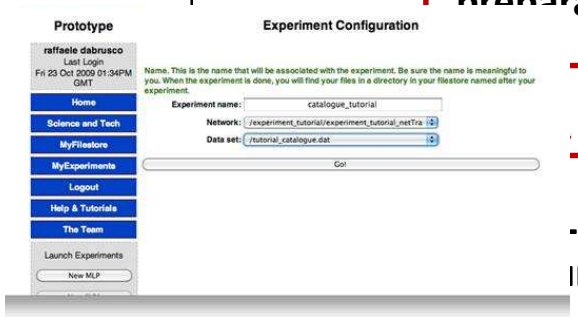
Test the network



Run the network

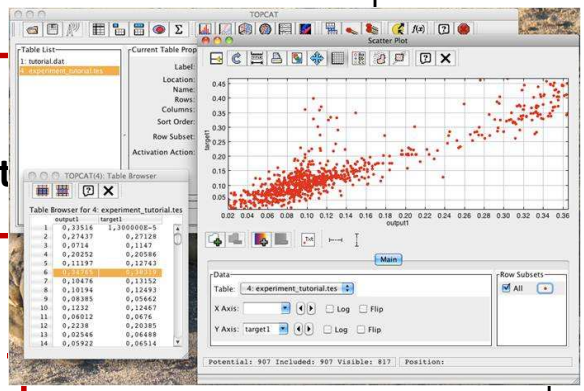


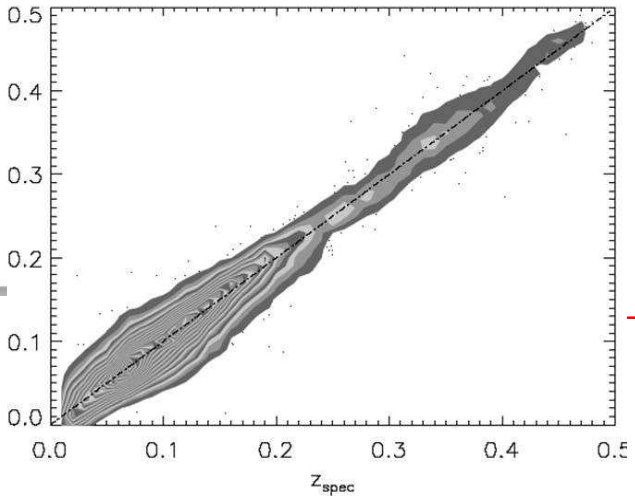
BOK



TRAINING

Nepo
reach





Run NN like a generic deterministic one-shot algorithm
Generate output Catalogue
Plot and analyze output

CATALOGUE

RUN

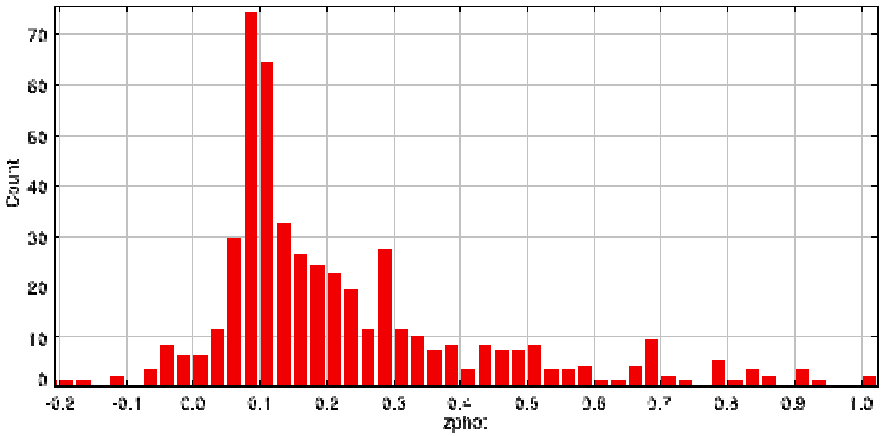
PLOT

Follow on-line tutorial

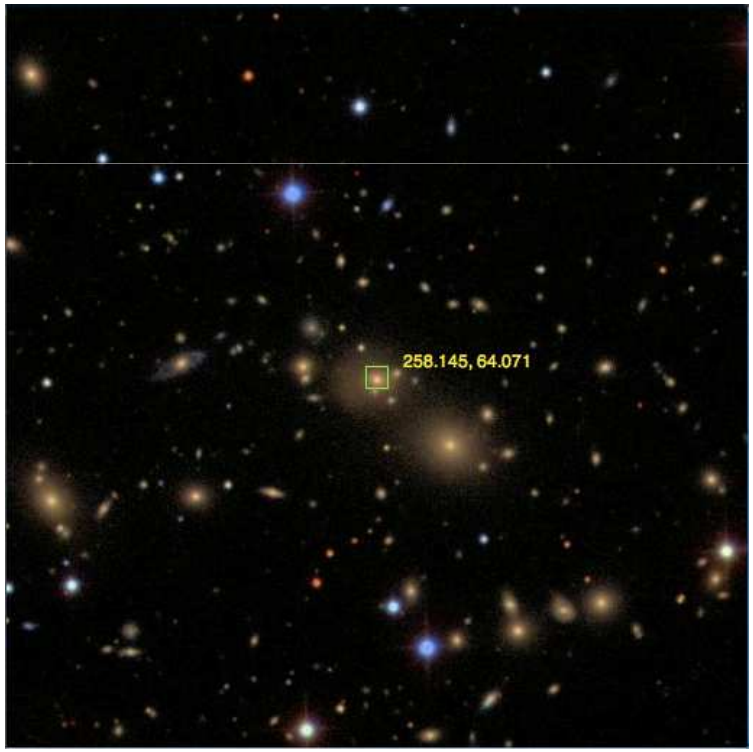
Use case II

A possible application of photometric redshifts is the selection of galaxies belonging to bound structures like clusters or groups (see, for example, [Capozzi et al. 2009]). Let's see how we can find out if a given observed "overdensity" of galaxies in a field is likely associated to a real structure.

Photometric redshifts can be used as a valid alternative technique or side-by-side to other methods based on the photometric properties of galaxies (for example, the "red-sequence" method [Gladders & Yee 2000]).



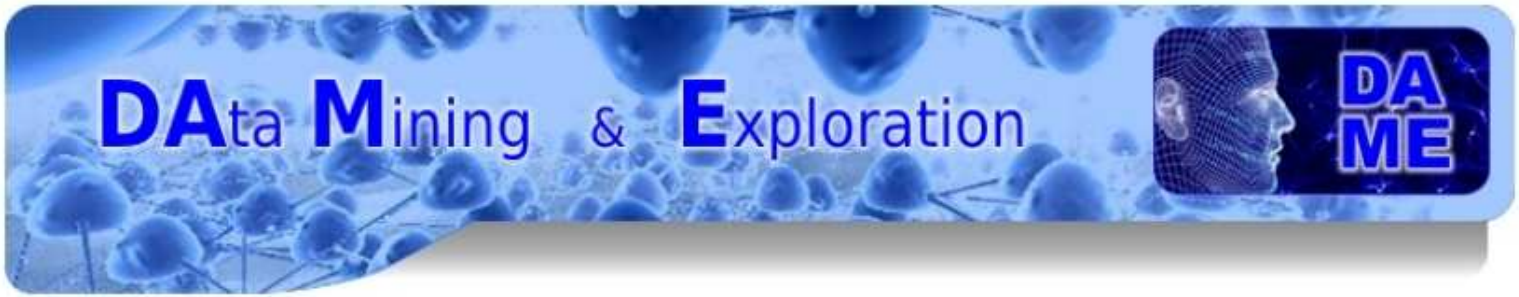
We will consider a **well known rich cluster of galaxies (Abell 2255)** and **explore a field containing this cluster using SDSS photometry and our own photometry redshifts**



You will use **DAME, Topcat and Aladin, with which you should already be comfortable**!

Tutorial – Part II

Follow the step-by-step instructions on the DAME prototype web page…



Prototype

Username
brescia
Password
Log in

Home
Science and Tech
Sign Up!
Help & Tutorials
The Team

Using photometric redshifts to spot galaxy clusters

A possible application of photometric redshifts relates to the selection of sources belonging to bound structures like clusters or groups of galaxies.

A close projected group of galaxies on the sky can be either produced by sources which are gravitationally connected to each other (i.e. physically close in both projected and redshift spaces), or can be the result of a chance superposition of physically unrelated galaxies (sources which are close in the sky projection but have different redshifts).

In order to check this out, **photometric redshifts can be used as a valid alternative technique or side-by-side to other classical methods** based on the photometric properties of galaxies (for example, the "red-sequence" method [Gladders & Yee 2000]).

Let's see how we can find out if a given "overdensity" of galaxies in a field is likely associated to a real structure. We will consider a **well known rich cluster of galaxies called Abell 2255** (after the name of the astronomer who compiled a large catalogue of clusters of galaxies) and **explore a field containing this cluster using photometry redshifts.**

We will **use DAME, Topcat and Aladin**, but only the main steps of the process will be described, so you'll have to work the details out by yourself!

1. First of all, **you need a catalogue of photometric sources** in the area of the sky where the cluster of galaxy is more likely to be found. We will use the querying service provided by Topcat to look, on Vizier, for sources in the Abell 2255 field. Click on the "Open new table" icon in the main bar of Topcat, then click on the Vizier service icon ; now you can perform a cone search around the Abell 2255 cluster position in the sky by filling the fields in the Vizier window with the following values: RA 258.145 (deg), DEC 64.071 (deg), Radius: 200" (change