DAME & NEXt – II

segmentation of images in a distributed computing environment

M. Brescia[1], F. Isgrò[2], O. Laurino[2], **G. Longo[2]**, C. Vellucci[2]
& the DAME/VONeural team at
S.Co.P.E. & Center for Advanced Computing Research CACR
(CALTECH)

1 – INAF Osservatorio Astronomico di Capodimonte
2 – Dipartimento di Scienze Fisiche Università Federico II di Napoli
3 - California Institute of Technology

SUMMARY OF THE TALK

- Astronomical and biomedical imaging
- Why distributed computing is needed
- Some applications in the DAME web application framework
- Issues to be solved for the porting to bio-medical applications
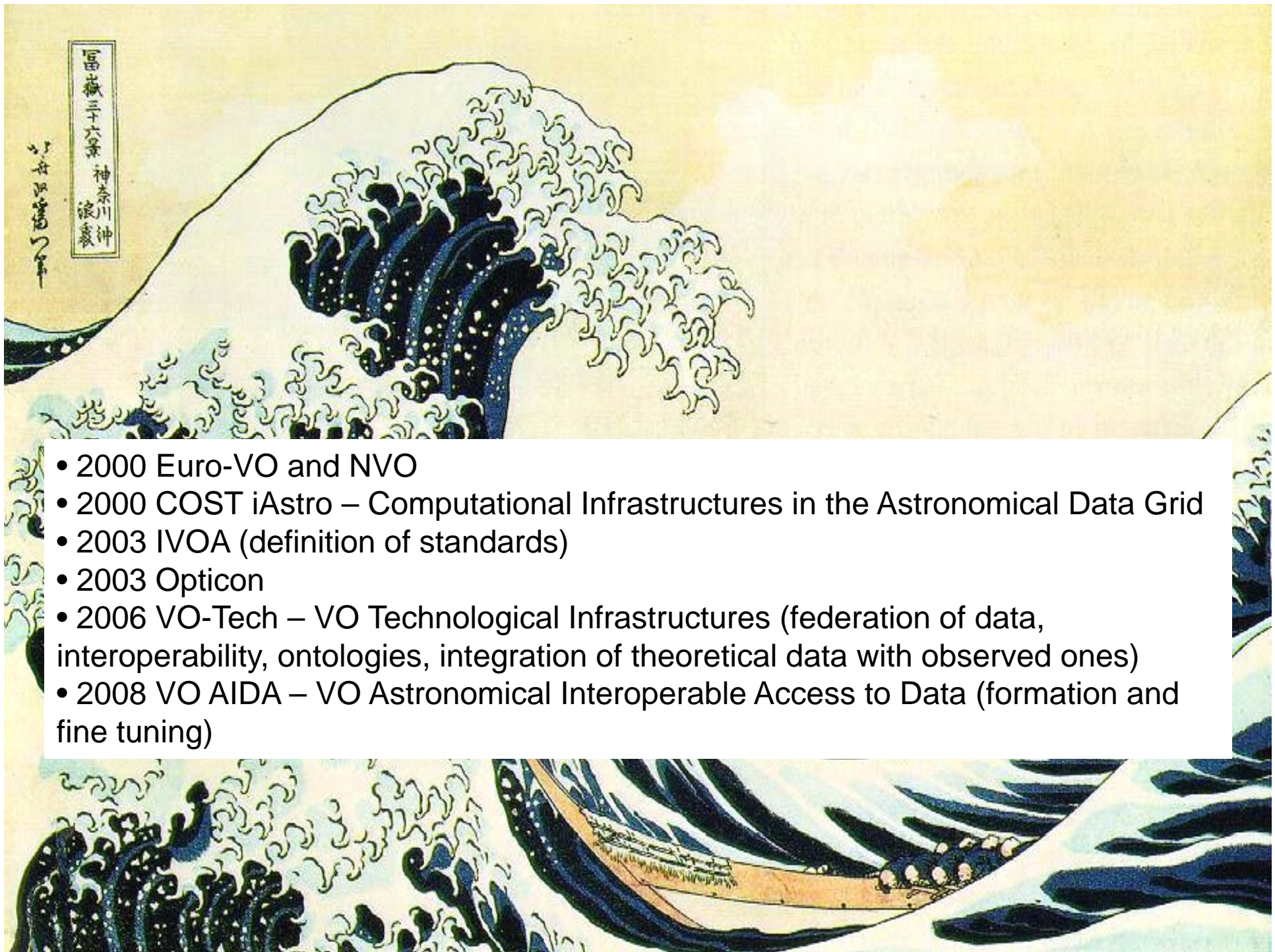
- 2000 Euro-VO and NVO
- 2000 COST iAstro – Computational Infrastructures in the Astronomical Data Grid
- 2003 IVOA (definition of standards)
- 2003 Opticon
- 2006 VO-Tech – VO Technological Infrastructures (federation of data, interoperability, ontologies, integration of theoretical data with observed ones)
- 2008 VO AIDA – VO Astronomical Interoperable Access to Data (formation and fine tuning)

Why an astronomer is here……

Astronomical images processing

- Large amount of archival data (many PBytes) distributed over three continents federated and interoperable (IVOA standards)
- Need to federate heterogeneous data (textual, tables, images) etc.
- Services to be accessed by a large community
- Heterogeneous data; proprietary (restricted access) and public data
- Real time processing requirement, large number of operations per pixel

Issues

- Need for the user to upload (in databases and to the pipeline storage) proprietary data
- Need to keep, update and access distributed archives
- Both amount of data and number of users impose automatic image processing methods

**DAME (Data Mining and Exploration)
a joint venture of S.Co.P.E. and Caltech
To offer as web applications a wide collection of
DM and DE methods and toos
(running on the GRID) to the scientific community**

DAME: project requirements
http://dame.na.infn.it/

1. Fully modular
2. Simple tasks are run on server
3. Computing intensive algorithms run on the GRID without need for personal certificates
4. NB: must not depend on the computing infrastructure: GRID, Cloud, etc.) where it is deployed.
5. Minimal data transfer (VO-space = storage element)
6. Possibility to build specific workflows and to insert user's modules as plug-in's
7. Advanced visualization capabilities (comsumed as web services)
8. Each module must be consumable as web service
9. Plastification for software interoperability
10. Security requirements

11. USER FRIENDLINESS!!!! (tutorials, on line manuals, high level technical docs, template use cases, etc…). USERS ARE NOT EXPERTS IN MATHEMATICS; COMPUTER SCIENCE; ETC)
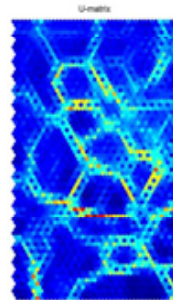
DAME – DAta Mining and Exploration
California Institute of Technology - Università degli Studi Federico II

What is DAME

DAME is a web application to perform data mining on massive data sets. In order to ensure scalability it allows the user to access distributed computing facilities provided by the Center for Advanced Research in Computing at Caltech and by the **S.Co.P.E.** project at the University of Napoli Federico II. DAME is derived from the **VO-Neural** project.

As a function of the size and complexity of your task, your computation will be re-directed to larger computing facility.

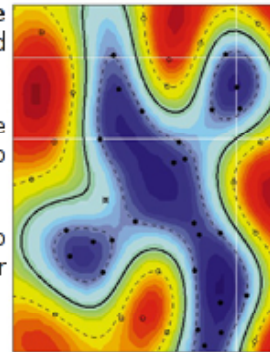**DAME** is an evolving platform. Therefore please provide us with your comments and feedbacks.

Start signing up for a **new account**. Signing up will provide you with a **persistant filestore** on our servers, so that you won't need to upload your datasets each time you want to perform a new calculation.
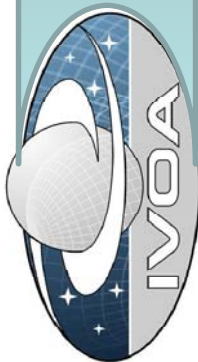
Your filestore will also contain all the **output files from the experiments you launch**, so that you can visualize or download them when the experiment is done.

During an experiment you can **visualize the log file** showing the status of the experiment and visualize output files. You can also **abort a calculation**.

You can even **download an entire directory** in a compressed zip archive on your hard disk. Output files can be used as inputs for other experiments, and so on...

In the **"Help & Tutorials" section** you will find **documentation, examples and tutorials**. The first time you login, your filestore will contain some datasets you can use following the tutorials.

DAME – DAta Mining and Exploration
California Institute of Technology - Università degli Studi Federico II

Giuseppe Iongo
Last Login
Mon 23 Feb 2009
01:06PM GMT

Home

MyFilestore

MyExperiments

Logout

Help & Tutorials

The Team

Links

Launch Experiments

New MLP

New SVM

New PhotoZ

Ministero degli

Experiment Details

Experiment Name: MAE	
Launched	
Parameter	**Value**
Data Set	/longo/pro/catalogue_photozs.csv

Dirs	Files	Actions
/longo/MAE		Download
	MAE.log	Delete
	catalogue_photozs.csv	Delete

Experiment Log
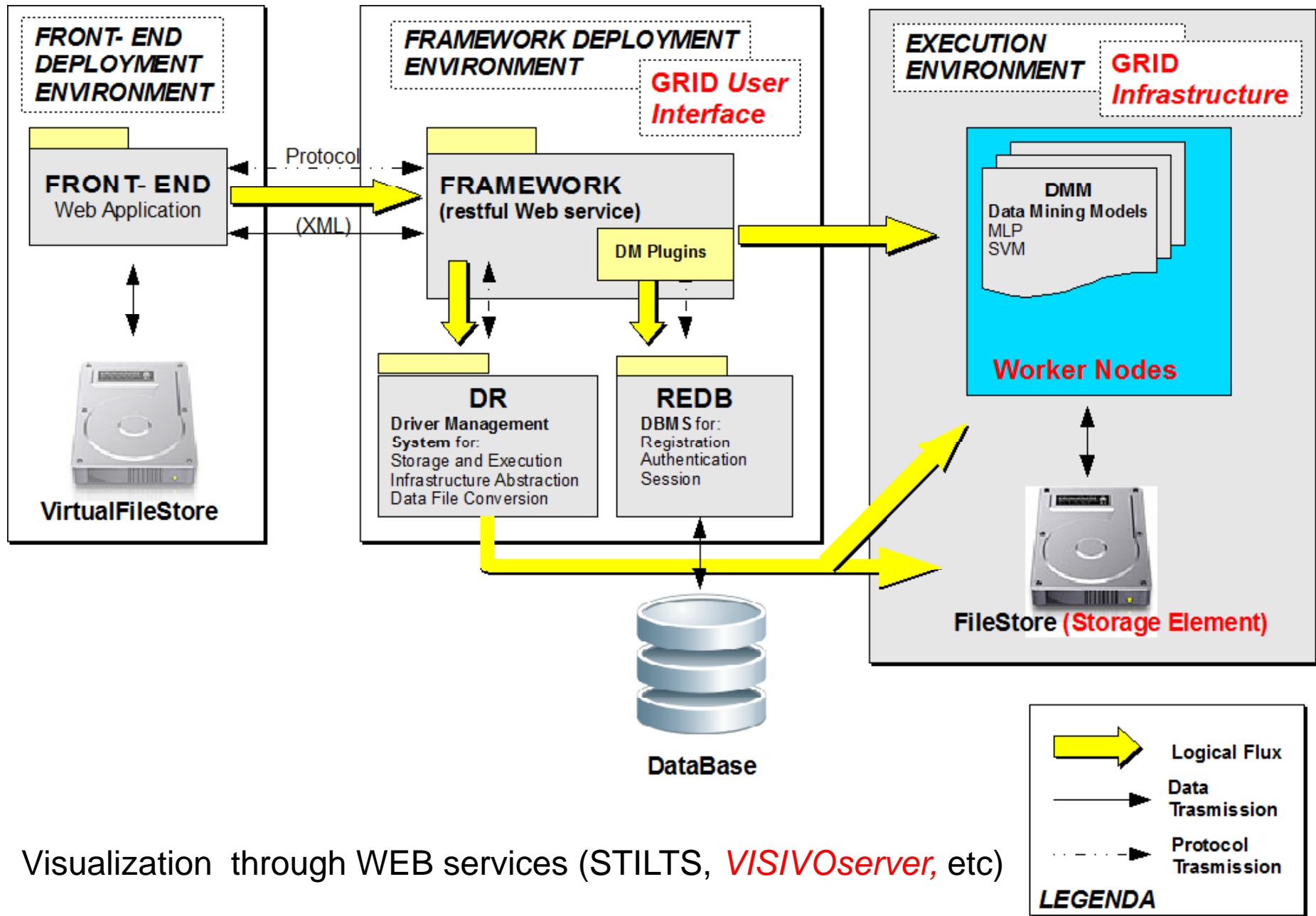
```
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Starting the evaluation of photometric redshifts!
Assigning photometric sources to clusters in the parameter space:
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Changing format of the intermediate files:
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
---------------------------------------------------------------------------------
Invoked csv2fits...
Reading input in .catalogue_photozs.csv
Written output in .catalogue_photozs.fits
Do not need to be moved...
End of csv2fits!
---------------------------------------------------------------------------------
```
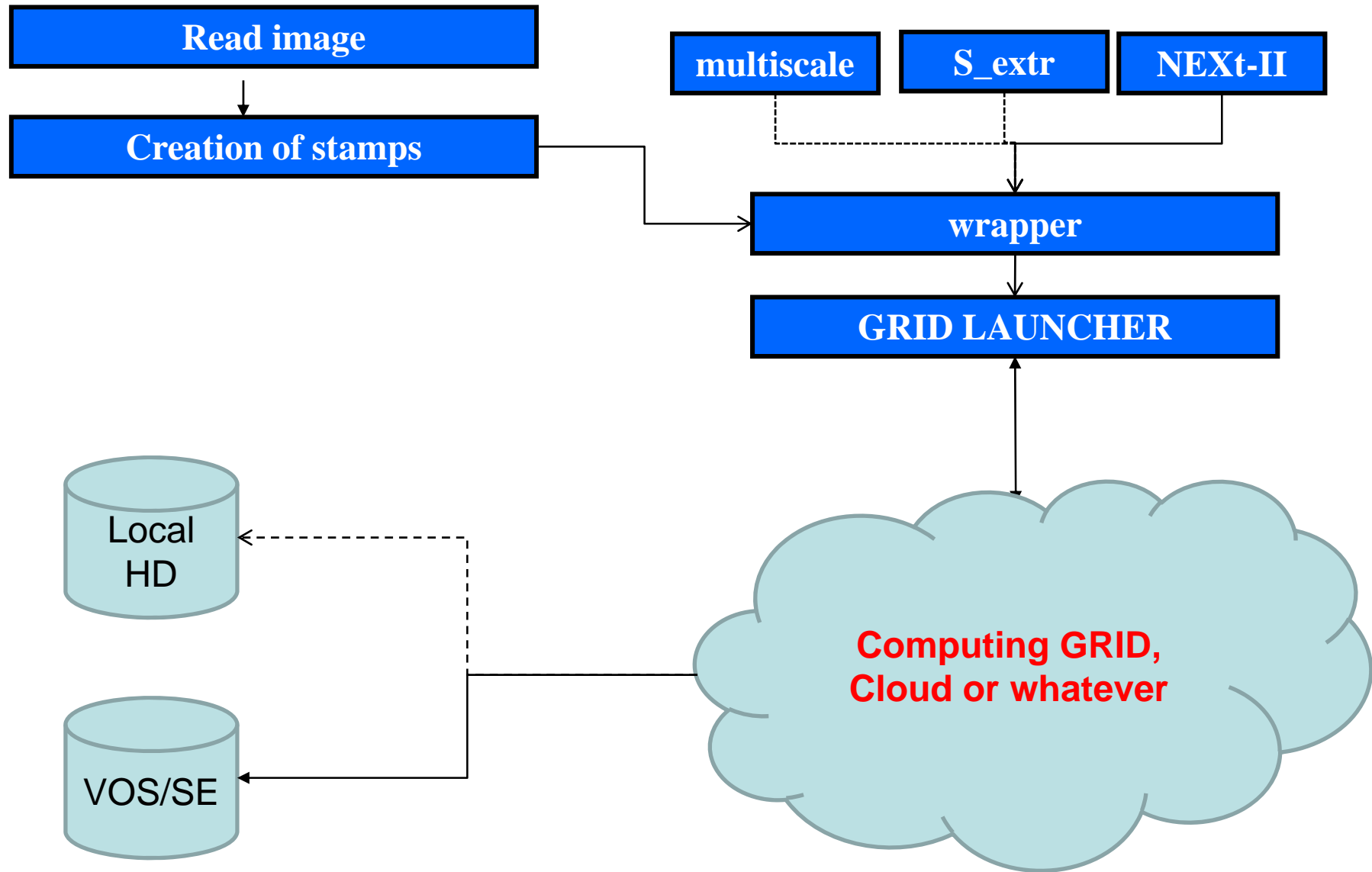
Plots

DB1　DB2　DB3　　　　DBn

VObs

Local HD

VO SPace

SE/VOS

DAME Front-end

DAME framework

DAME-DMM

| NEXT-II | SVM | MLPs | SOM | Mn |

Results/Visualization
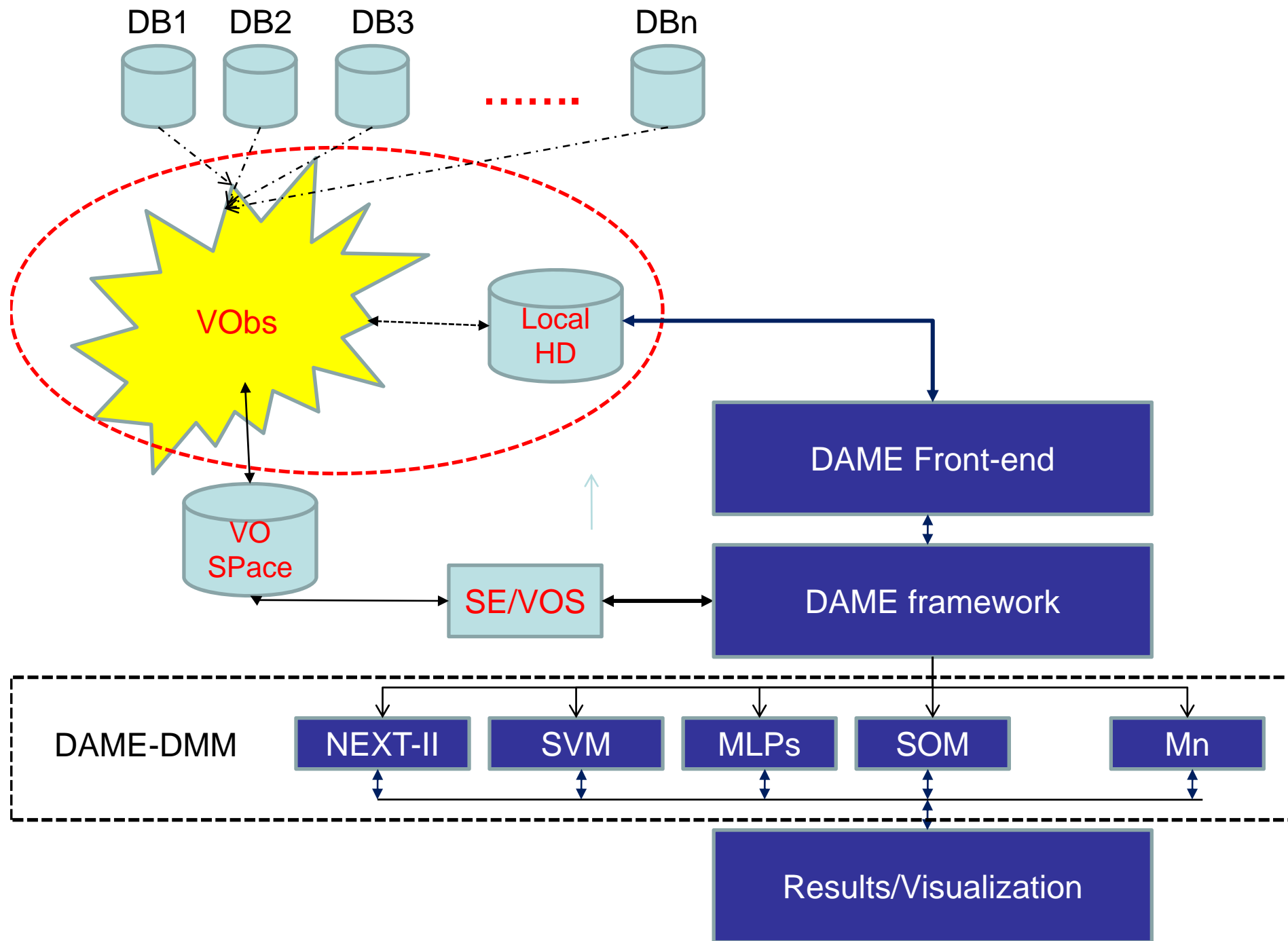
Visualization through WEB services (STILTS, *VISIVOserver,* etc)

FRONT- END DEPLOYMENT ENVIRONMENT

FRONT- END
Web Application

VirtualFileStore

FRAMEWORK DEPLOYMENT ENVIRONMENT

GRID *User Interface*

Protocol

(XML)

FRAMEWORK
(restful Web service)

DM Plugins

DR
Driver Management
System for:
Storage and Execution
Infrastructure Abstraction
Data File Conversion

REDB
DBMS for:
Registration
Authentication
Session

DataBase

EXECUTION ENVIRONMENT

GRID Infrastructure

DMM
Data Mining Models
MLP
SVM

Worker Nodes

FileStore **(Storage Element)**

LEGENDA

Logical Flux

Data Trasmission

Protocol Trasmission

NExt: Neural Extractor

A little history:

NEXt was conceived for the segmentation of astronomical images in signal/background pixels

- Astronomical images are highly anysotropic with low S/N ratio
- Modern Digital surveys produce volumes of data in the Pbyte domain
  (… larger than most biomedical images archives)
- Number of operations per pixel >100

Goals

- Identify withouth a priori assumptions pixels containing signals from objects against pixels affected by background only
- After classification to identify objects as connected regions of object pixels
- Create a "binary" mask containing the location of objects
- **Automatic procedure, stable and with no need for specific settings of the parameters**

Algoritmi user - dependent lead to non reproducible results
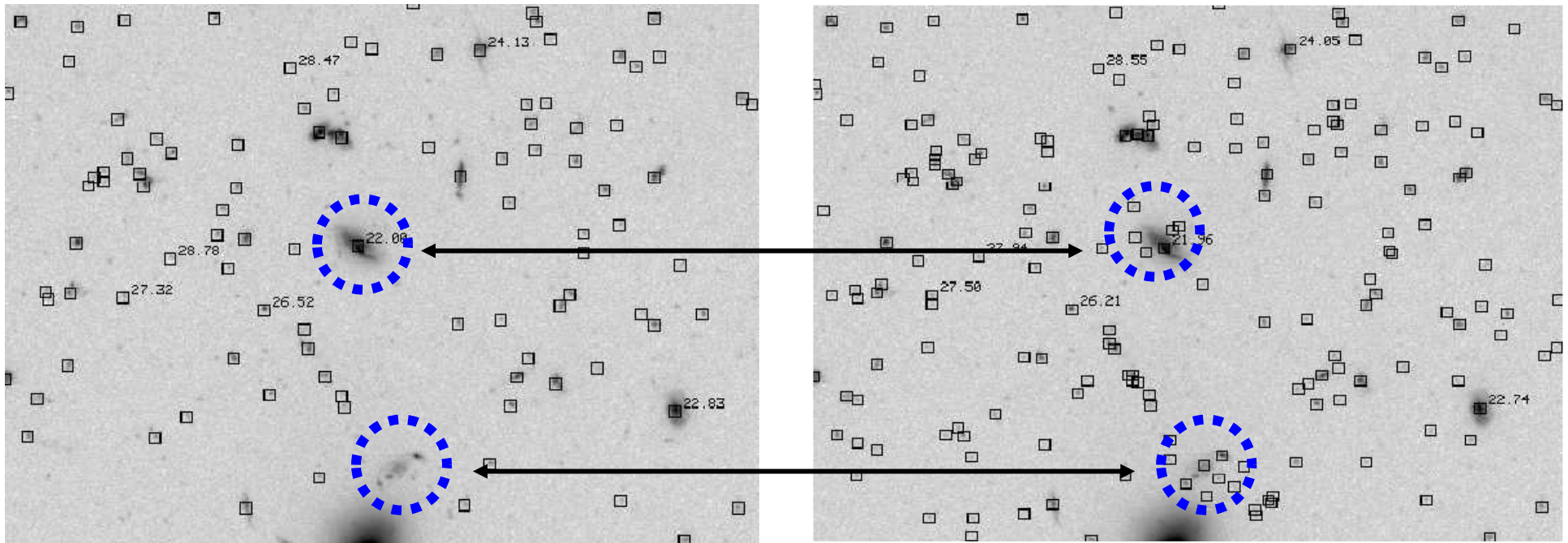
2 different runs of S-Extractor on the same field *(Ferguson et al. 1998)*

- Arbitrary detection treshold, user dependent parameters

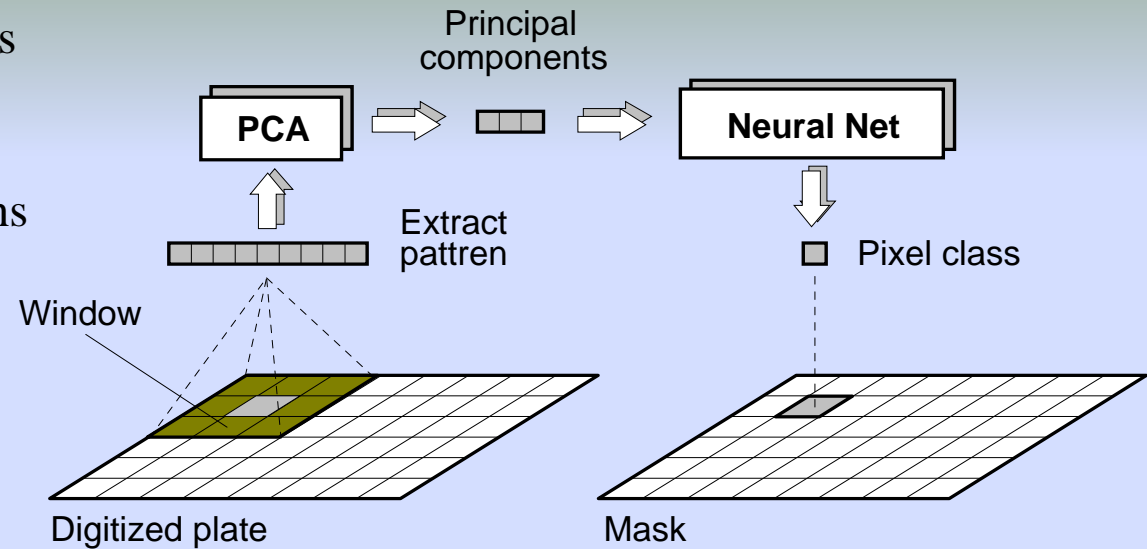- Catalogues are different expecially at low flux levels
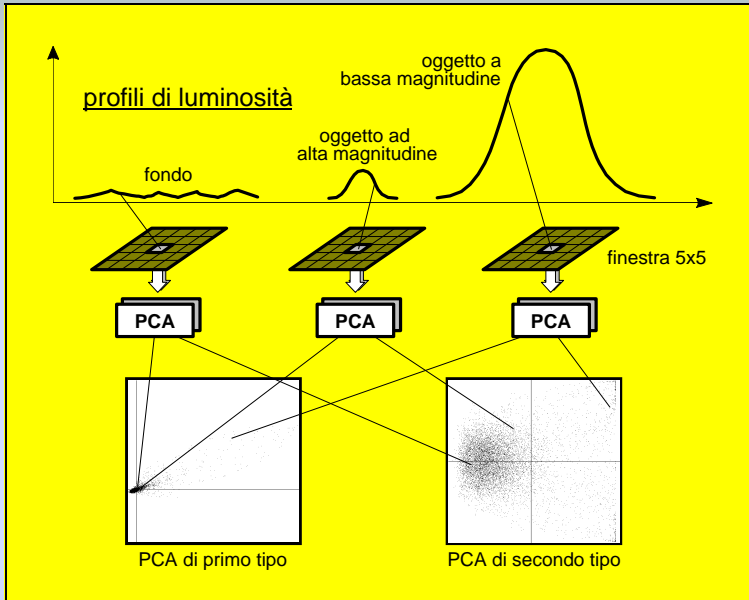
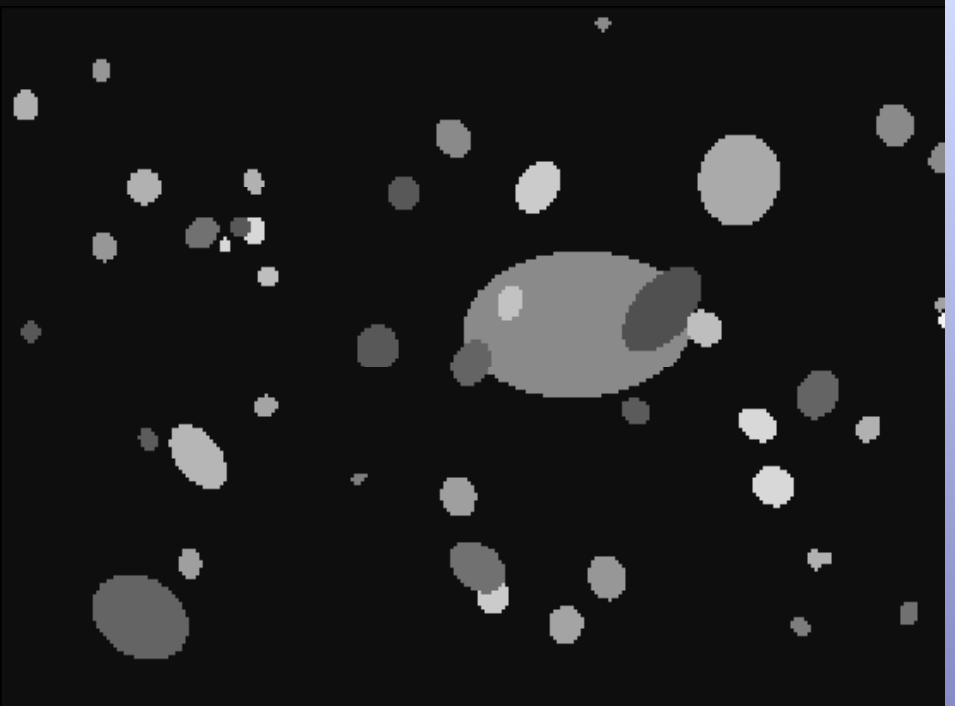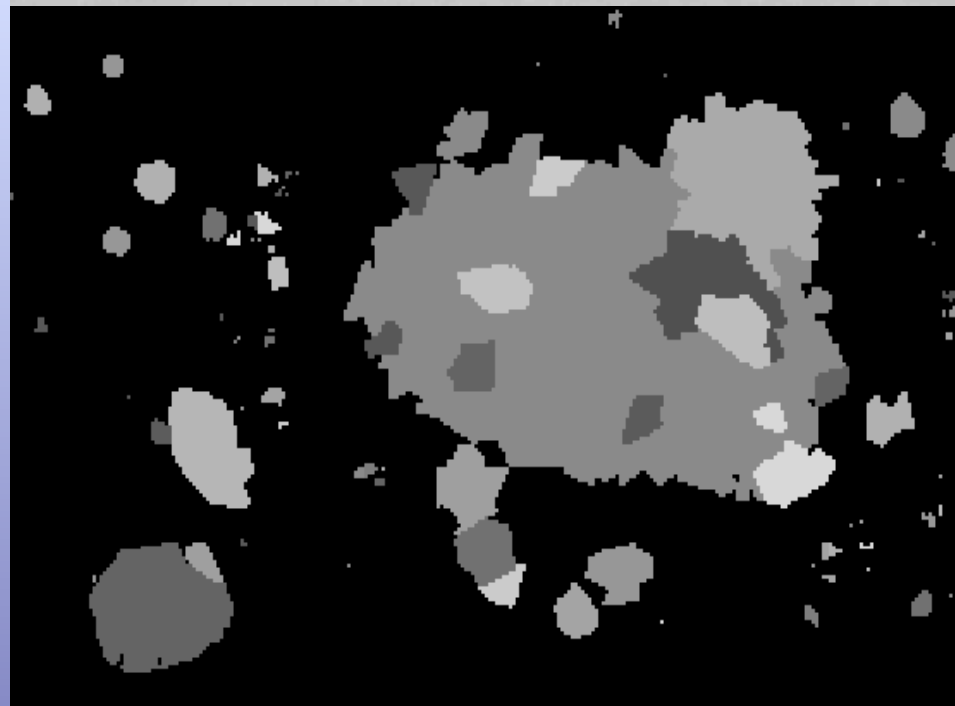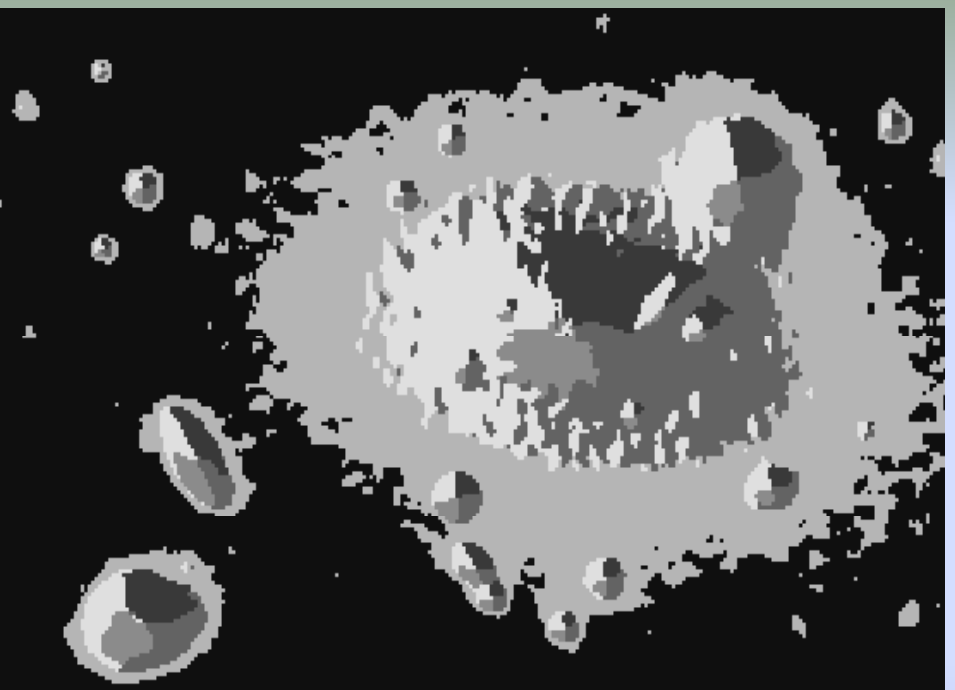**Necessità di funzioni di trasferimento (attivazione) non lineari**
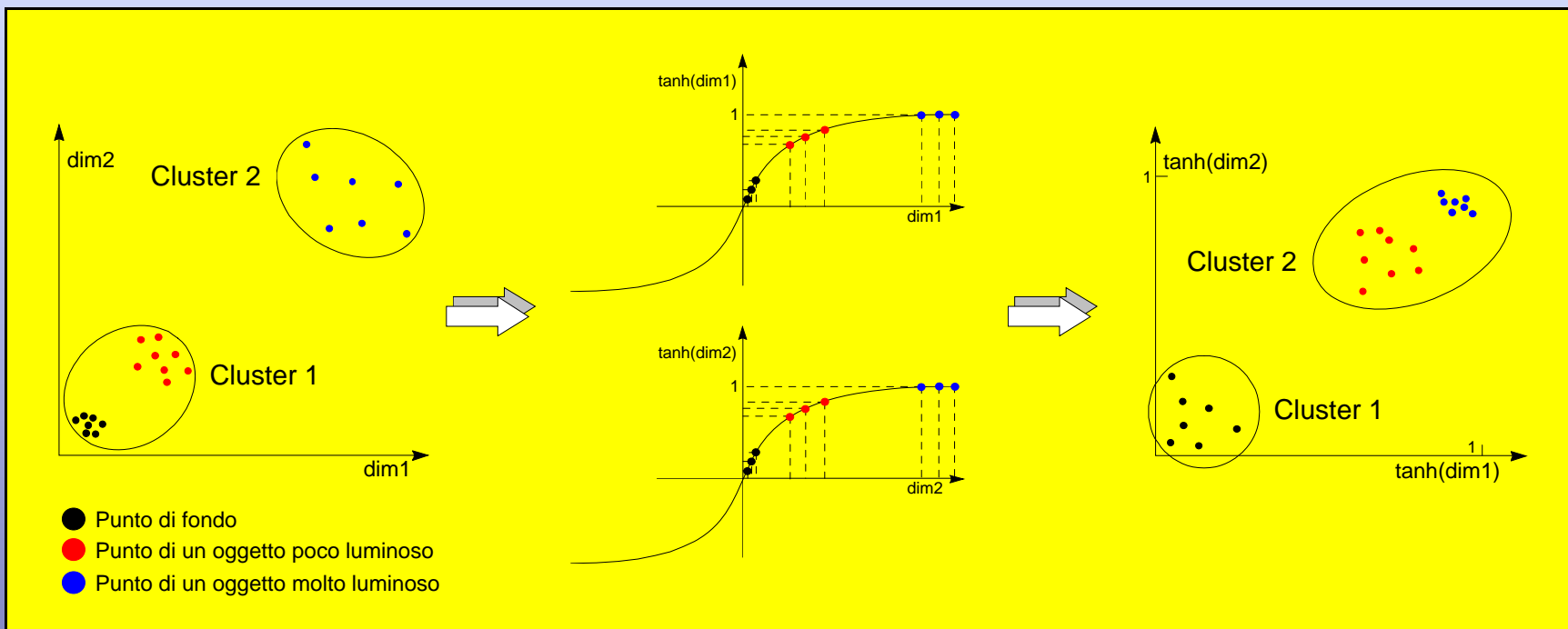
We Aimed To:

1) Manage simultaneously 3 bands

2) Improve detection correctness

3) Generate accurate classifications

Principal components

PCA

Neural Net

Extract pattren

Pixel class

Window

Digitized plate

Mask

Single band segmentation

Multiple band segmentation

Extract pattern

PCA

Principal components

Three-dimensional window

Neural Net

Pixel class

Plate digitized on 3 bands

Mask

Dall'alto verso il basso e da sininstra verso destra:

1) ML-SOM
2) ML-Neural Gas
3) ML-KMeans
4) ML-Maximum Entropy
5) GCS + ML-Neural Gas

6) Neural Gas
7) Neural Gas + Nearest Neighbour
8) Neural Gas + met. del Centroide
9) Neural Gas + media tra gruppi
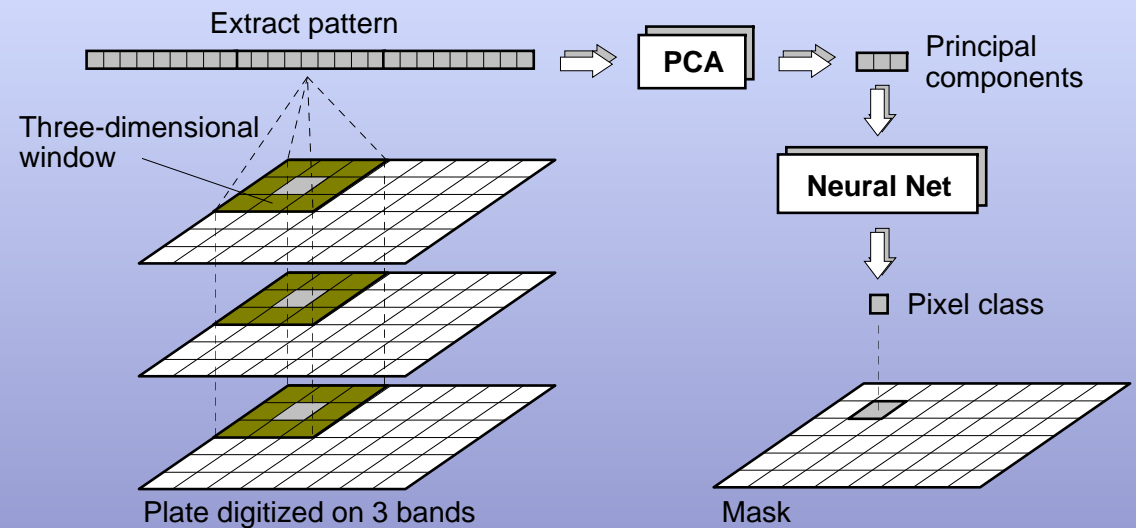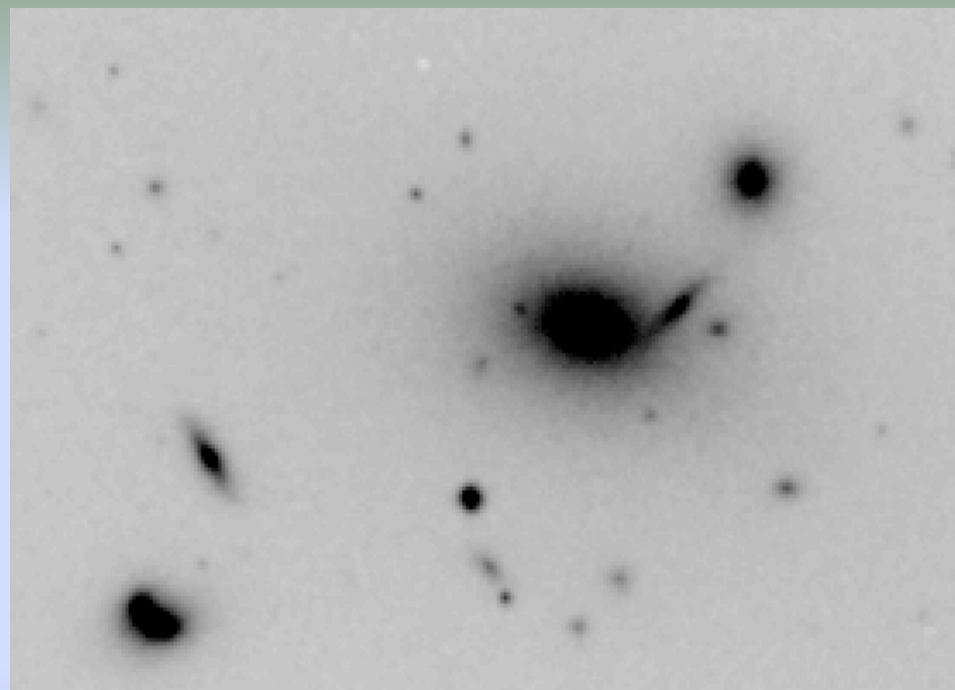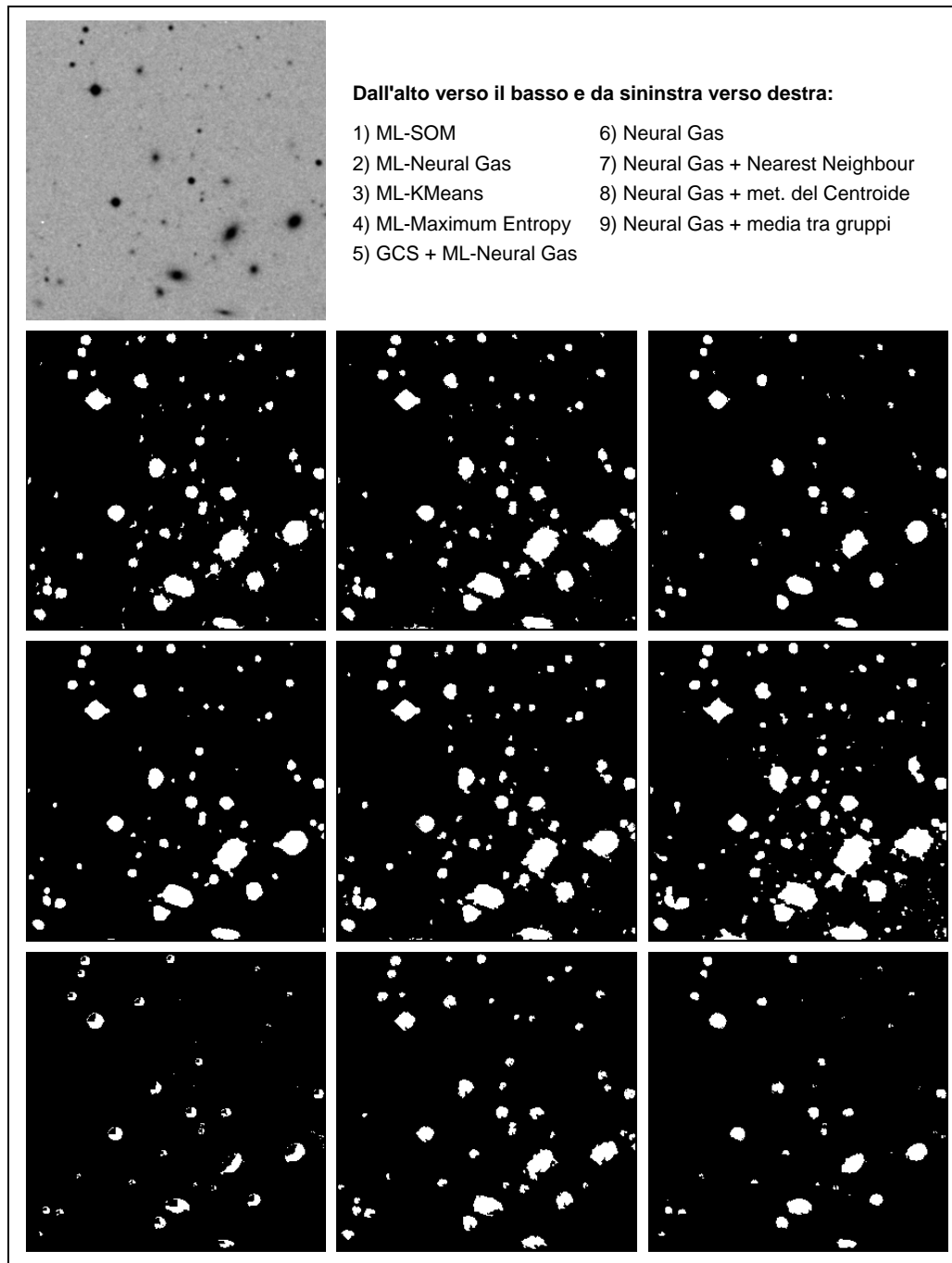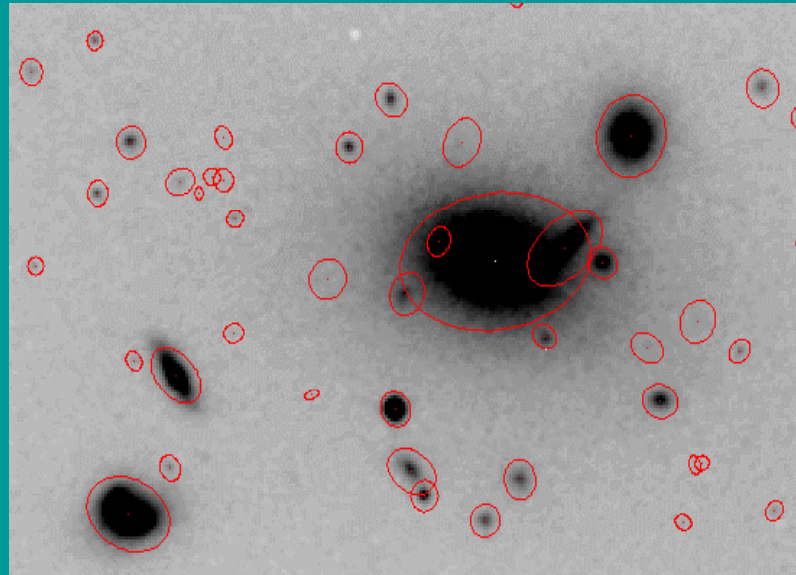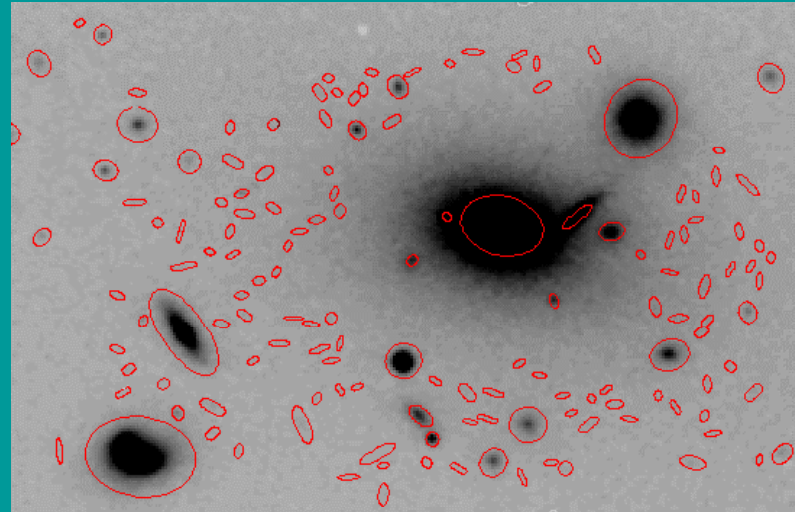
fig. 6.9 - esperimenti di segmentazione dell'immagine in alto.

NExt vs SKICAT

NExt after deblending   **SKICAT**

- La maggior parte degli oggetti è "vera"

- Decompone bene oggetti sovrapposti

- Necessità di un lungo post-processing dei cataloghi per pulirli

- Oggetti estesi sono spezzati in frammenti

- Diametri e fotometria errati

What comes NEXT

- Addition of other segmentation models
- Parallelization of some parts of the codes
- Porting and fine tuning to other domains
- PLASTIFICATION (or SAMPIFICATION) of methods