The computer does the hard work

## (semi)-automatic classification of zooplankton with ZooImage

Philippe Grosjean & Denis Kevin

Numerical Ecology of Aquatic Systems, Mons-Hainaut University, Belgium

'OPC to CCD' seeking for a brain

- Hadware is definitely available now to planktonologists as Mark just explained

- Gathering terabytes of plankton images is easy

- Transforming the pixels into biologically/ecologically meaningful data is the next challenge

  The most sophisticated oceanographic imaging system remains desperately useless without an equally sophisticated software to process the huge amount of raw data it produces
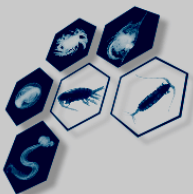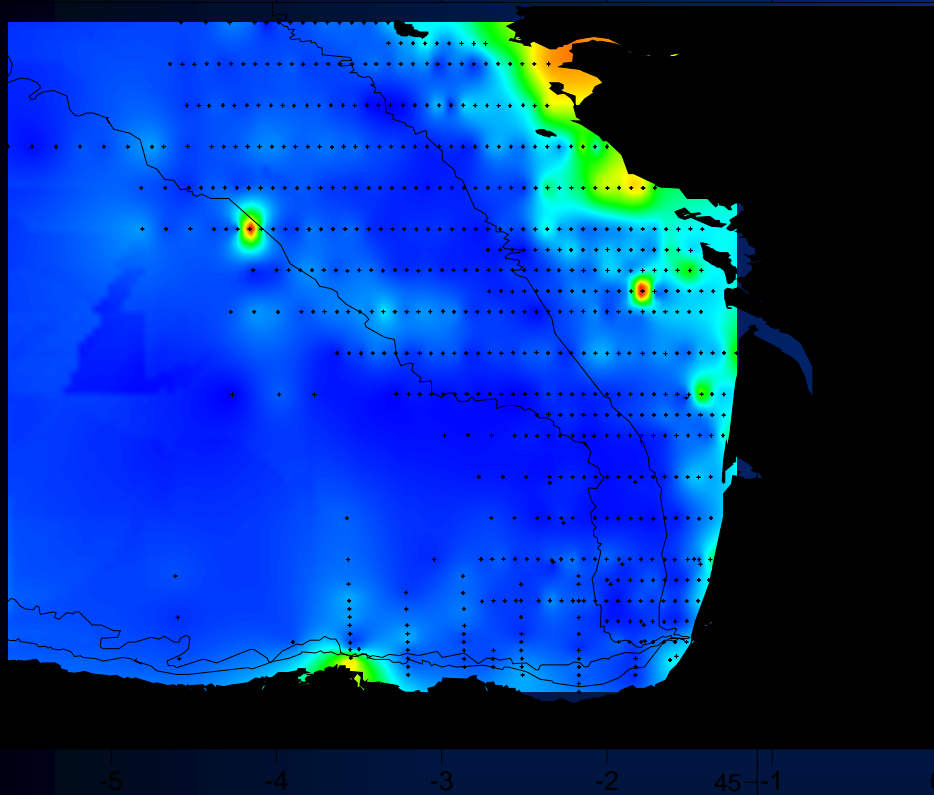
Oceanographer wants/needs more data

- Need for both (3D)-space and time coverage across large areas

- Patchiness of plankton distribution imposes to increase the sampling density

- Colorful maps are often more striking than simple plots; oceanographers love them, but maps are data-hungry beasts!
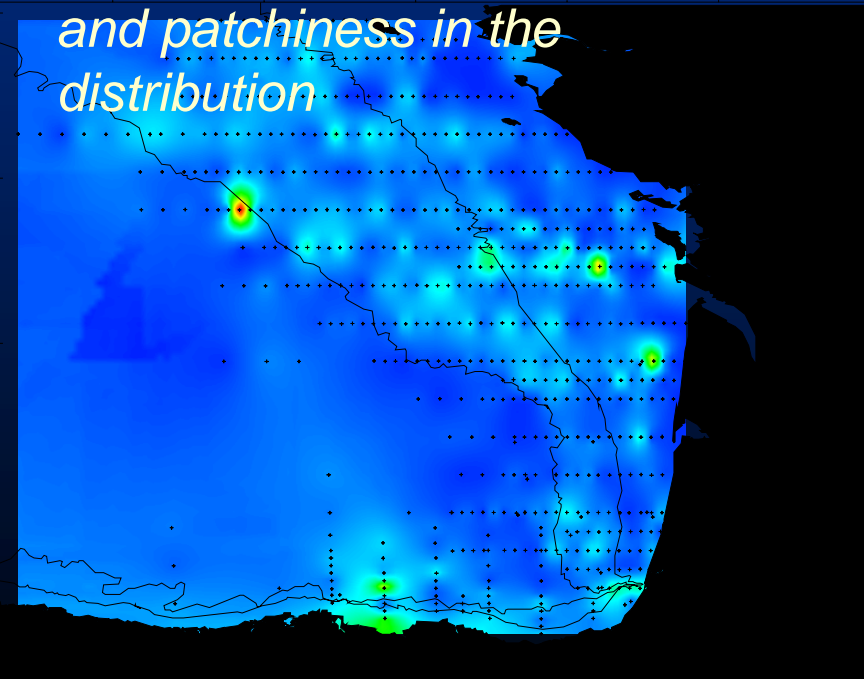
More data: These maps are much more appealing...

Zooplankton abundance in the Bay of Biscay at two different dates

Note the large number of stations sampled (black dots) and patchiness in the distribution
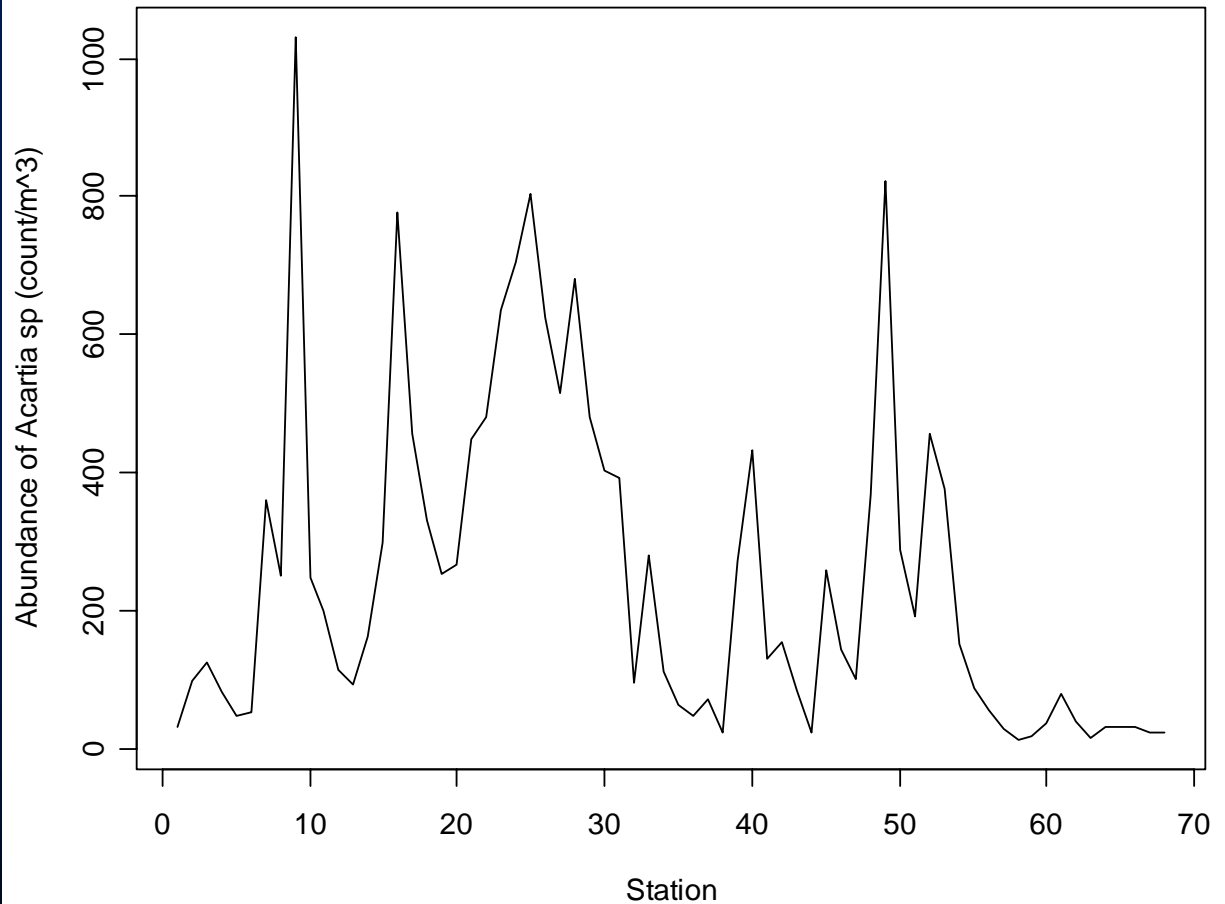
Data and graphs: X. Irigoien et al, AZTI. Analysis of plankton digital images with PVA, a precursor of ZooImage
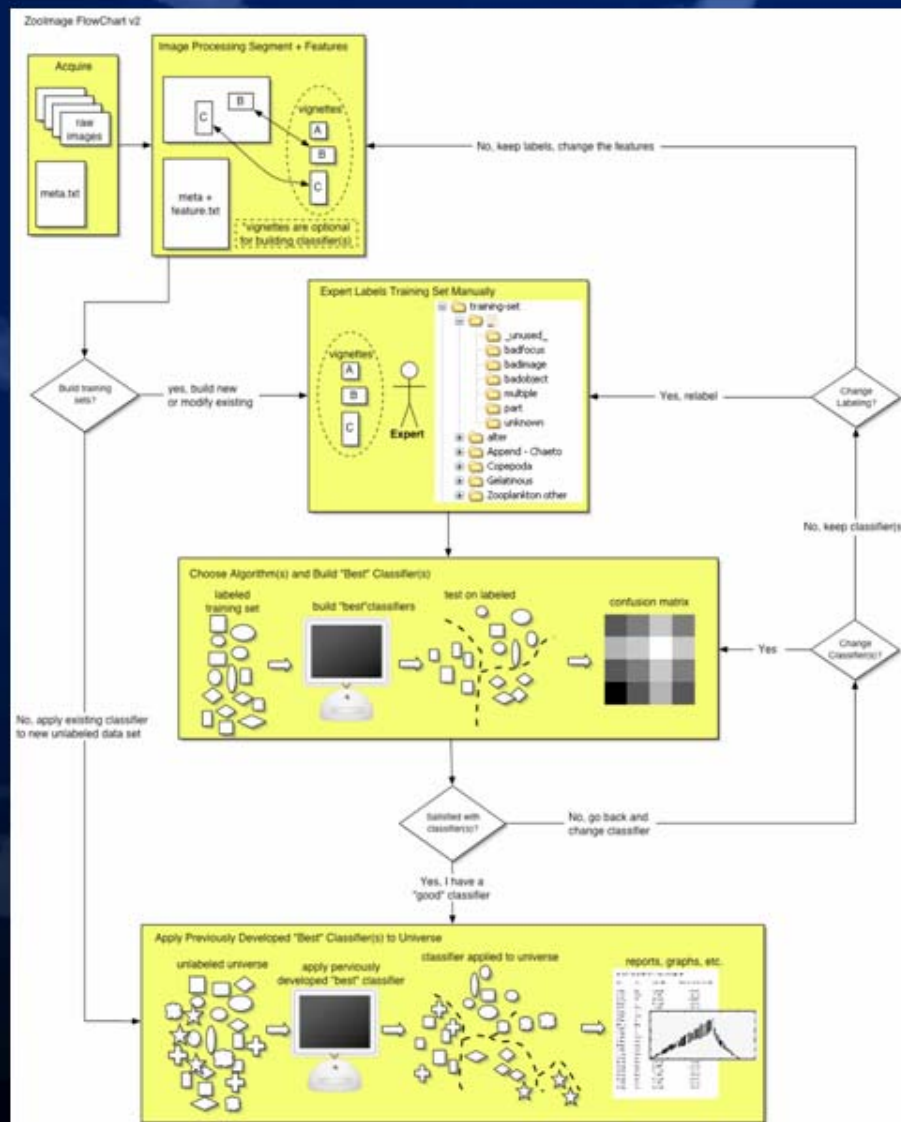
… than this plot

- Abundance of a copepod along a transect

- Space and time are confused

- No 2D/3D resolution

Plancton image analysis is complex

- **Image analysis (of zooplankton) is a complex task**

- **It requires a complex and specialized software**

- **Example: data process flowchart in ZooImage**

(diagram by Ben Tupper)

Plancton image analysis is complex

ge analysis (of
lankton) is a
plex task

- It requires a complex and specialized software

- Example: data process flowchart in ZooImage

(diagram by Ben Tupper)

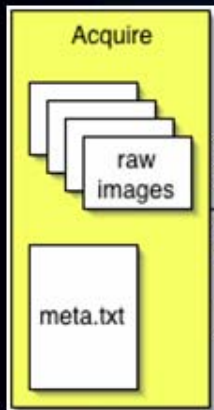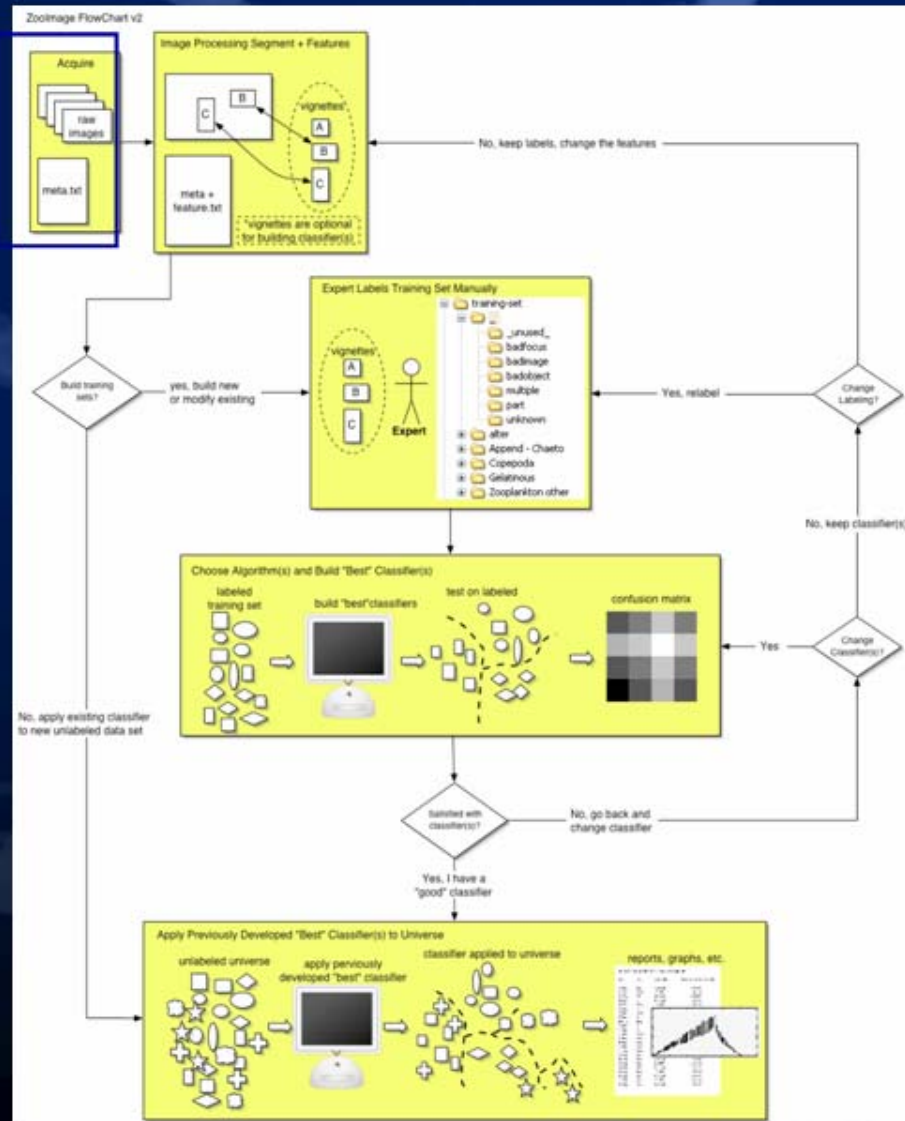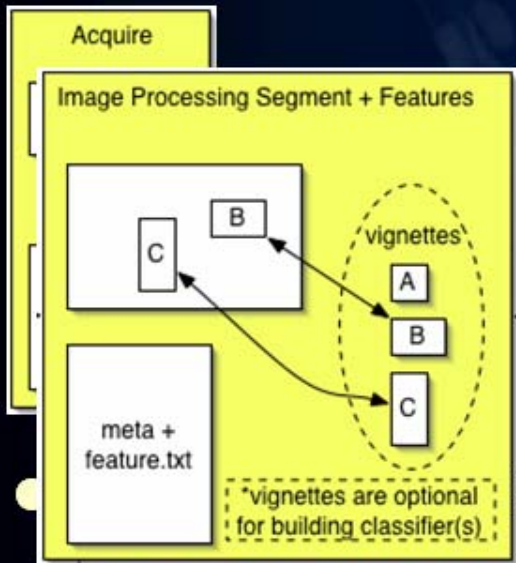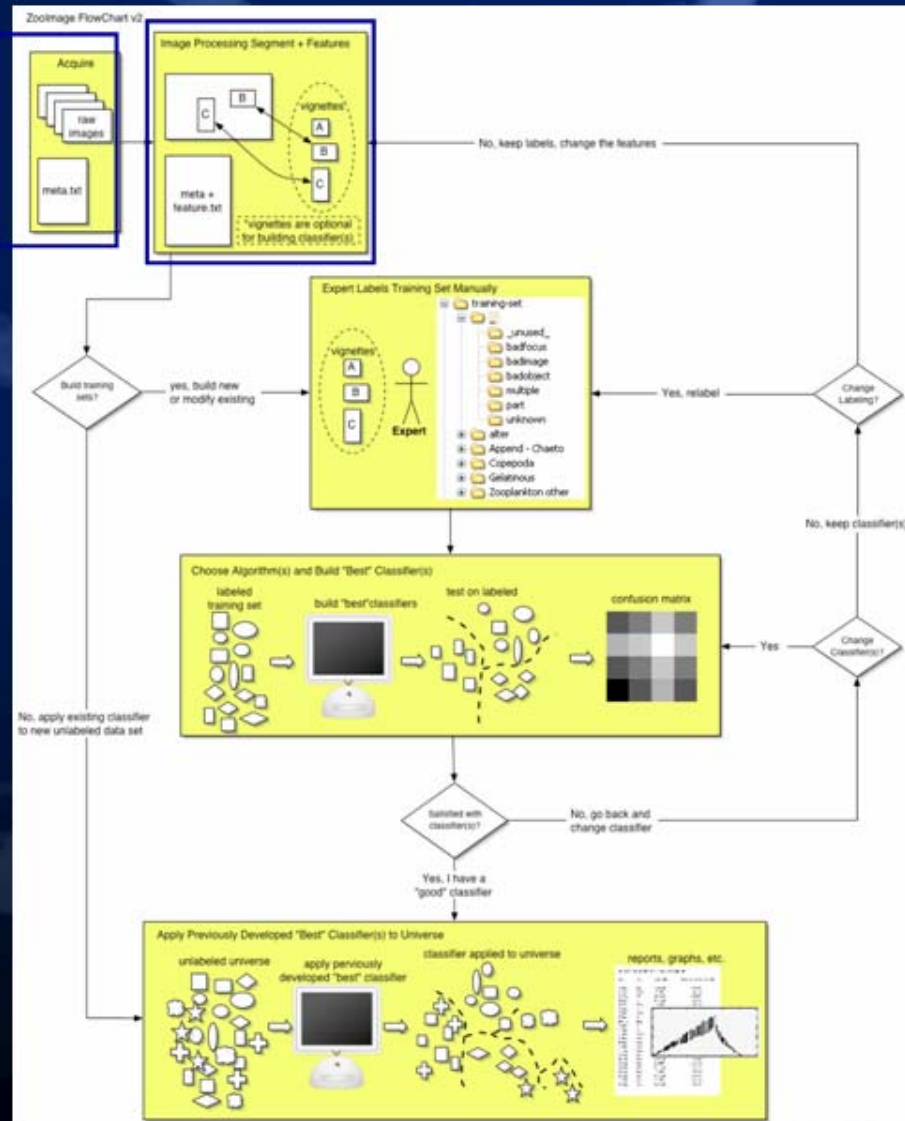# Plancton image analysis is complex

- ... is (of ... is a ...
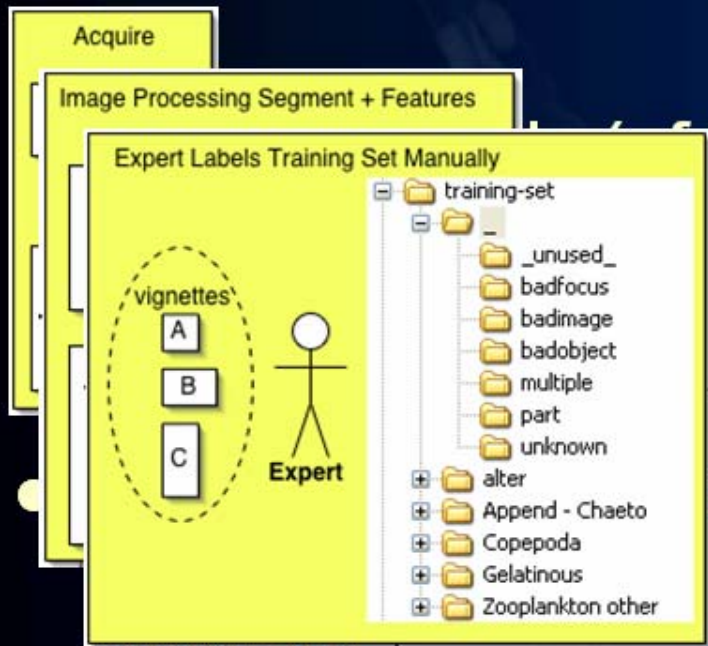- ... complex and specialized software

- Example: data process flowchart in ZooImage
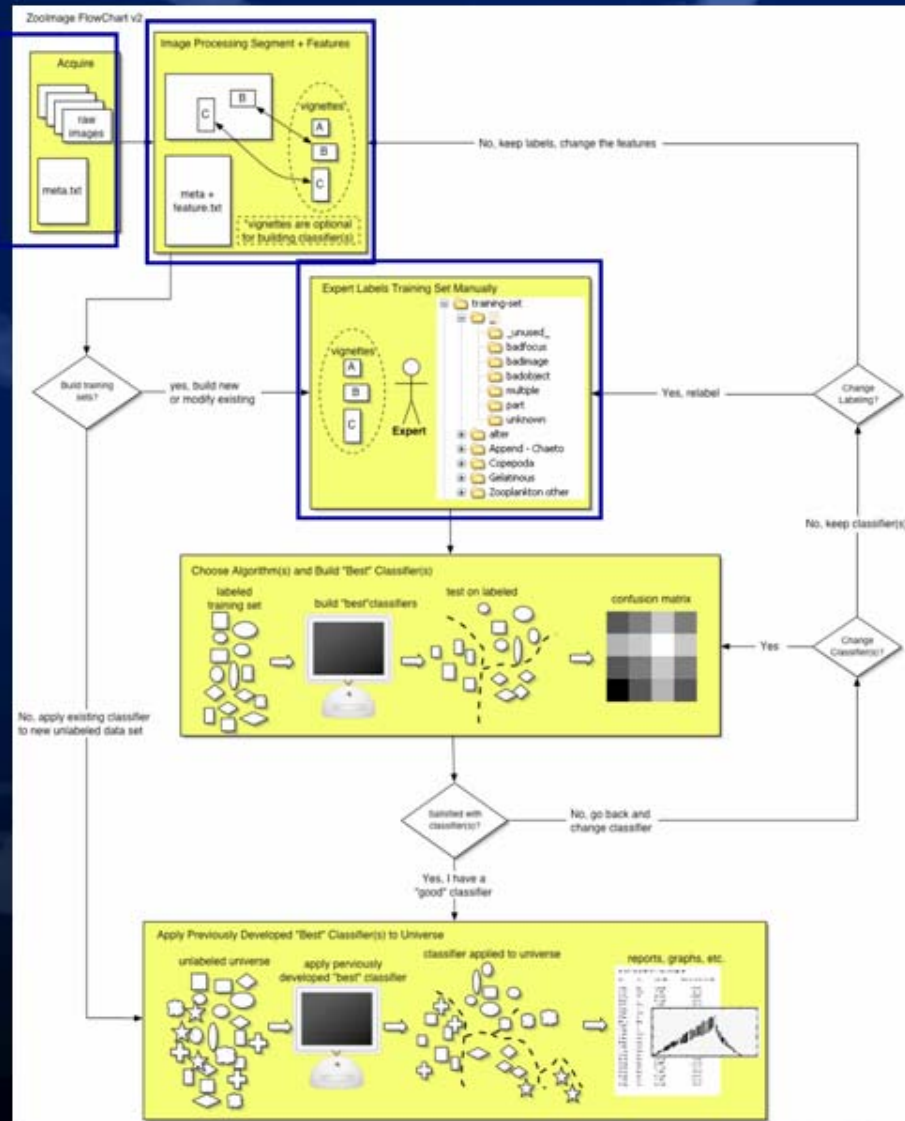
(diagram by Ben Tupper)
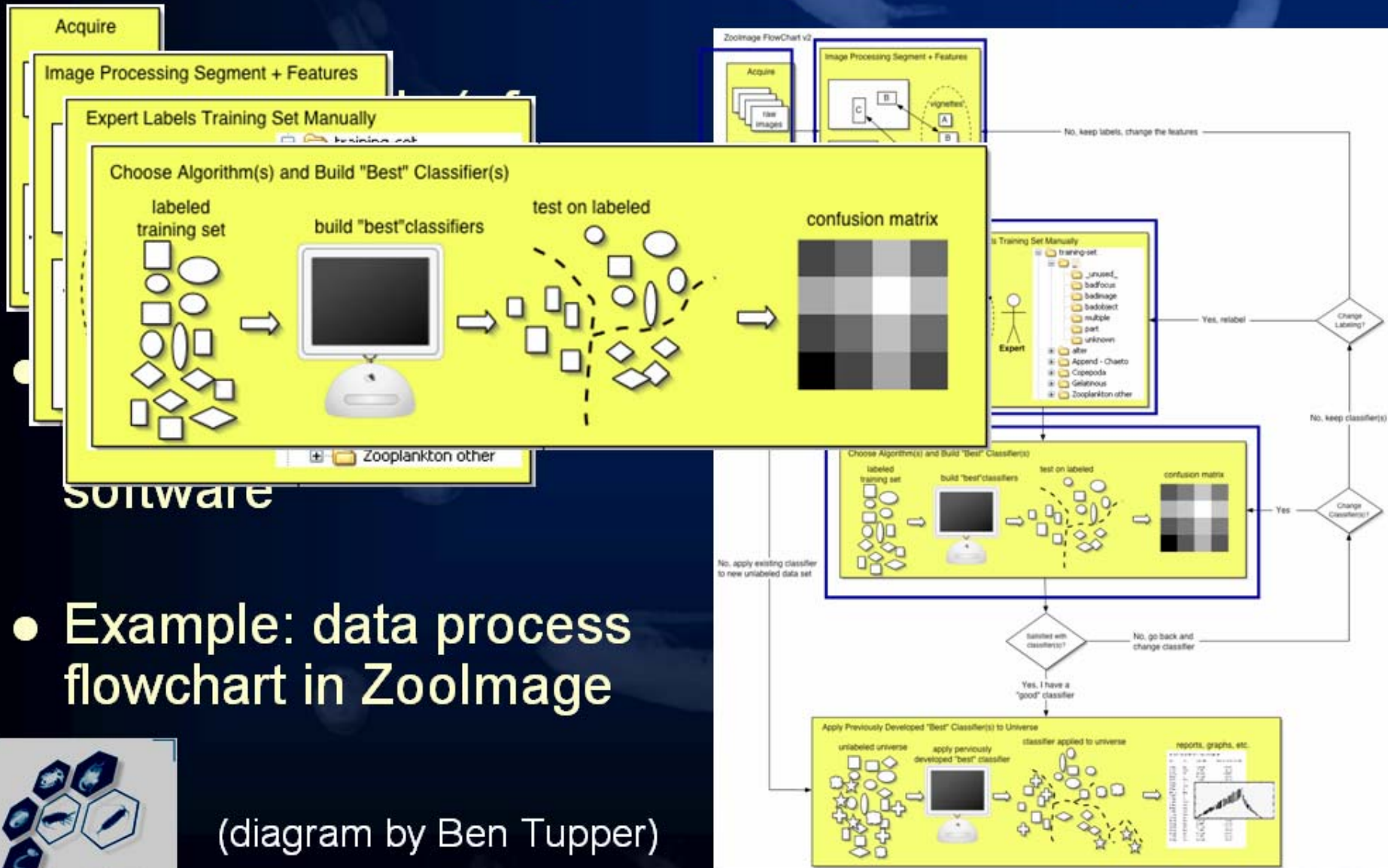
Plancton image analysis is complex

- ...software

- Example: data process flowchart in ZooImage

(diagram by Ben Tupper)

Plancton image analysis is complex

- Example: data process flowchart in ZooImage

(diagram by Ben Tupper)

Plancton image analysis is complex

Acquire

Image Processing Segment + Features

Expert Labels Training Set Manually

Choose Algorithm(s) and Build "Best" Classifier(s)

Apply Previously Developed "Best" Classifier(s) to Universe

unlabeled universe   apply perviously developed "best" classifier   classifier applied to universe   reports, graphs, etc.
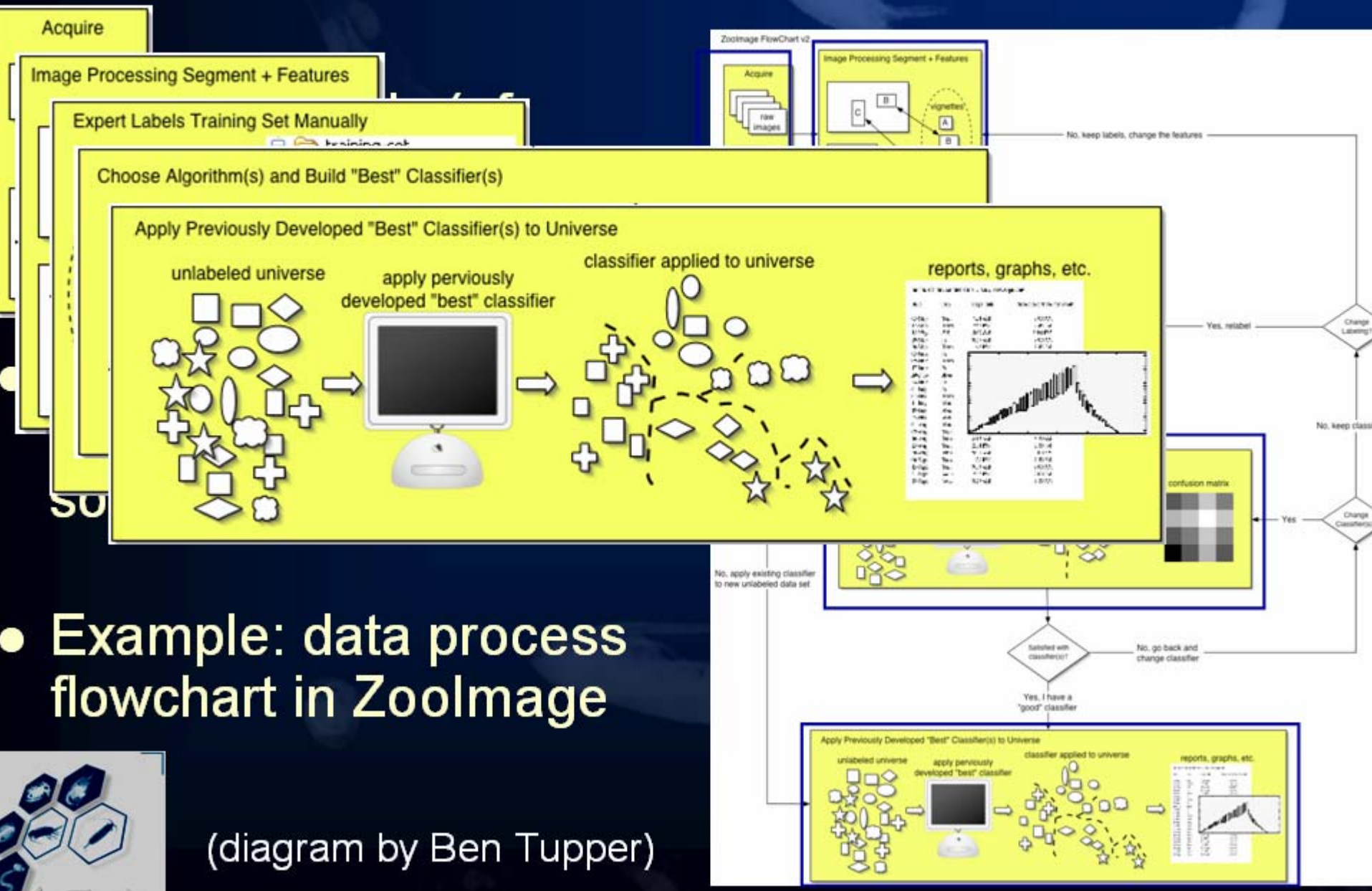
- Example: data process flowchart in ZooImage

(diagram by Ben Tupper)

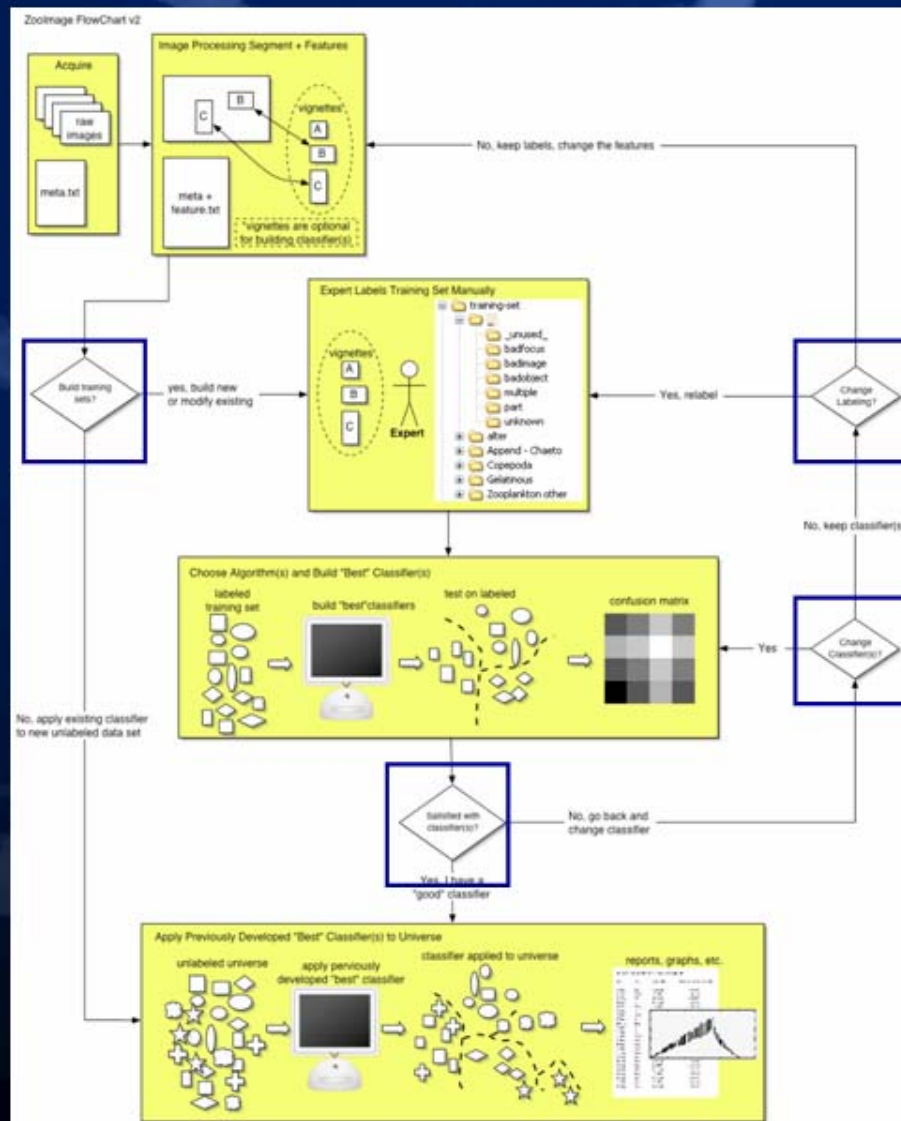# Plancton image analysis is complex

- Image analysis (of zooplankton) is a complex task

- It requires a complex and specialized software

- Example: data process flowchart in ZooImage

(diagram by Ben Tupper)

Why software developement lies behind hardware development?

- **Academic:** The 'politically correct' way of working does not favor software tools development (scripts ⇨ data analysis ⇨ results ⇨ publication(s) ⇨ 'don't-waste-you-time-and-think-at-your-next-publication').

- **Business:** The hardware sells better than the software. Complex software such as for plankton image analysis costs too much to develop, for a too small market to be profitable
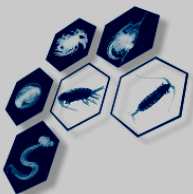
How to do then? (theory)

- Collaborate on a common 'Open Source' project (avoid reinventing the weel continuously)

- Provide a flexible and general framework
  - Single-use scripts of 'academics' become easily plugins shared with the community

  - Commercial solutions of 'businessmen' are leveraged by a free software to analyze the data…
    « hey! Good news: the brain exists already, and it is free. Let's just implement the bridge between our
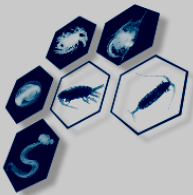
How to do then? (practice)

- Write and donate to the community a starting point for the Open Source software. Ex: Visual Plankton, ZooProcess (see corresponding posters), ZooImage

- Convince developers and users (ZooImage workshop, San Sebastian, November 2005)

- Convince managers it is worth financing such kind of project: Benfield et al, 2007, « white paper » to be published in Journal of Oceanography

- Organize international collaboration: SCOR Working Group 130. First meeting just after the symposium

The ZooImage software and framework

http://www.sciviews.org/zooimage

ZooImage history and current state

- **2004-2005:** first routines at UMH and PVA at AZTI

- **November 2005:** First public release of ZooImage at the GLOBEC SPACC ZooImage workshop (San Sebastian, Spain), for developers

- **November 2006:** version 1.0-0, features freeze, beta version

- **current:** version 1.2-1, usable by <u>experienced</u> data analysts, stability, 'friendliness' and end user's documentation to be perfected/completed for universal use.

- Windows plateform today, Linux and Mac OS ports in the future.

… see contributions (6) in session 9 for actual use
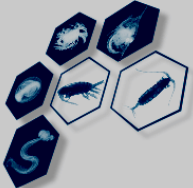
ZooImage is a collaborative work

- **Xabier Irigoien et al**, AZTI: PVA, workshop ZooImage, San Sebastian, november 2005 (also with **Angel Lopez-Urrutia**)

- **Mike Sieracki & Ben Tupper**, Bigelow Lab: FlowCAM plugin

- **Nick Mortimer**, CSIRO: PlanktonJ (see corresponding poster)

- **Jens Rasmussen**: forum (http://zooimage.overchord.net)

… more in the future !?

Plankton Image analysis forum: exchange between developers & users
http://zooimage.overchord.net

- Started in November 2005 and has grown steadily

- General topics and application specific areas for support and discussion

- Activity monitoring shows increasing interest among our community

http://zooimage.overchord.net
Unique Visists

http://zooimage.overchord.net
Visitor hits
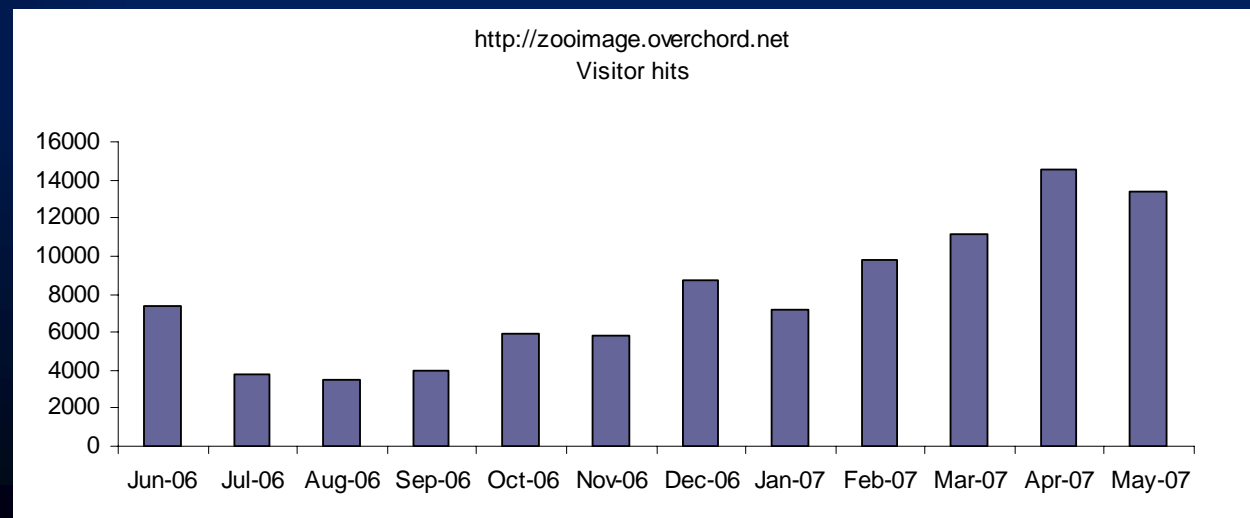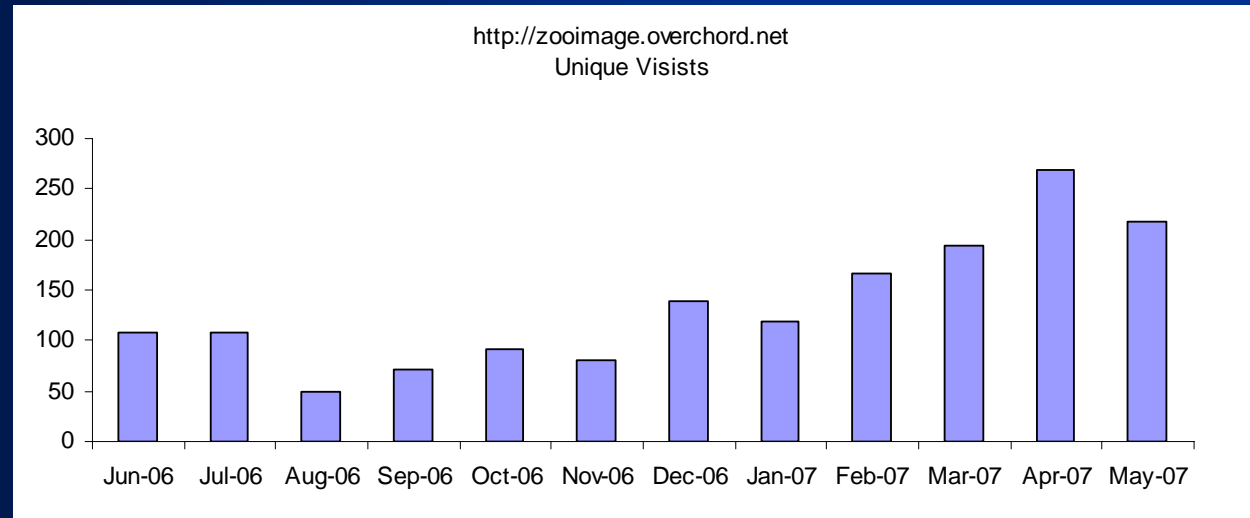
ZooImage framework (I)

Images of various origins → **import** → Common format ← **document** ← Metadata
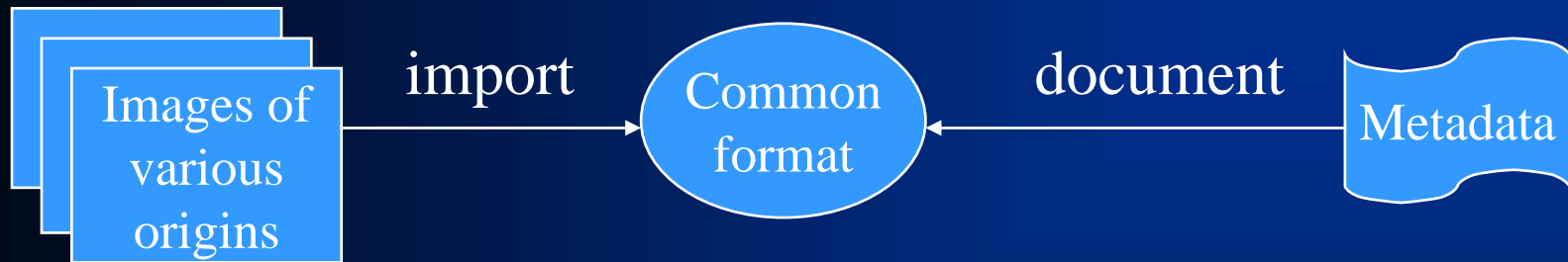
- **Standard naming conventions** for image files: SCS[.xxxx-xx-xx].SS+F[nn] eases exchange of data.

- **Standard metadata fields** ease automation of tasks.

```
BIO.2000-05-05.p72+A.zim - Sc1
File  Edit  Search  View  Tools  Options  Language  Help
ZI1
[Image]
Author=Naiara_Serrano
Hardware=HP_Scanjet_8200
Software=HP_driver_version_2.0
ImageType=reflect_24bits_color_600dpi

[Fraction]
Code=A
Min=-1
Max=-1

[Subsample]
SubPart=0.024
SubMethod=volumetry
CellPart=1.00
Replicates=1
VolIni=10.56
VolPrec=
```

Flexibility: image *versus* sample

Sample → **fractionning** → Fraction B / Fraction A → **digitization** → Images (replicates)
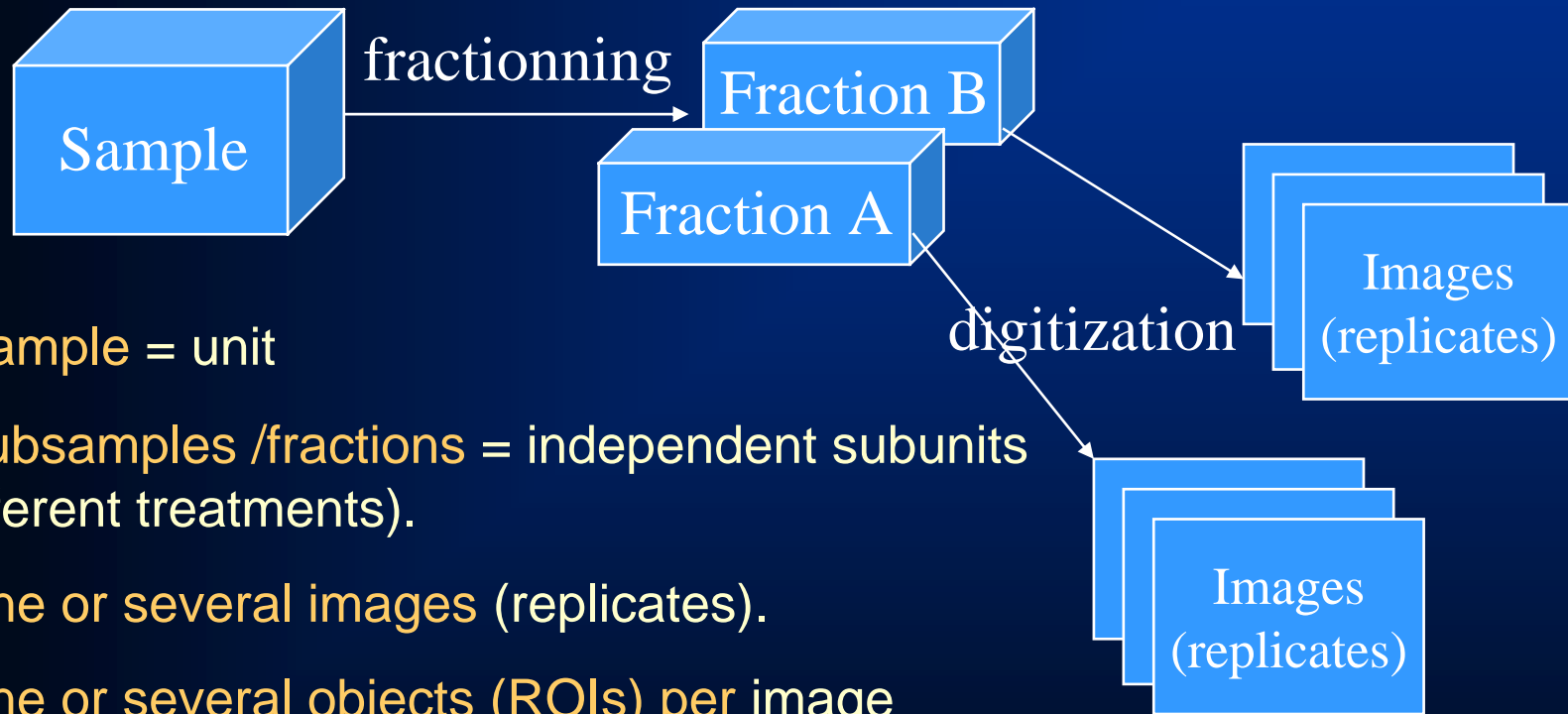
- Sample = unit
- Subsamples /fractions = independent subunits (different treatments).
- One or several images (replicates).
- One or several objects (ROIs) per image

=> *Most schemes are taken into account*

Separate plugins for image processing

Images of various origins → **import** → Common format ← **document** ← Metadata
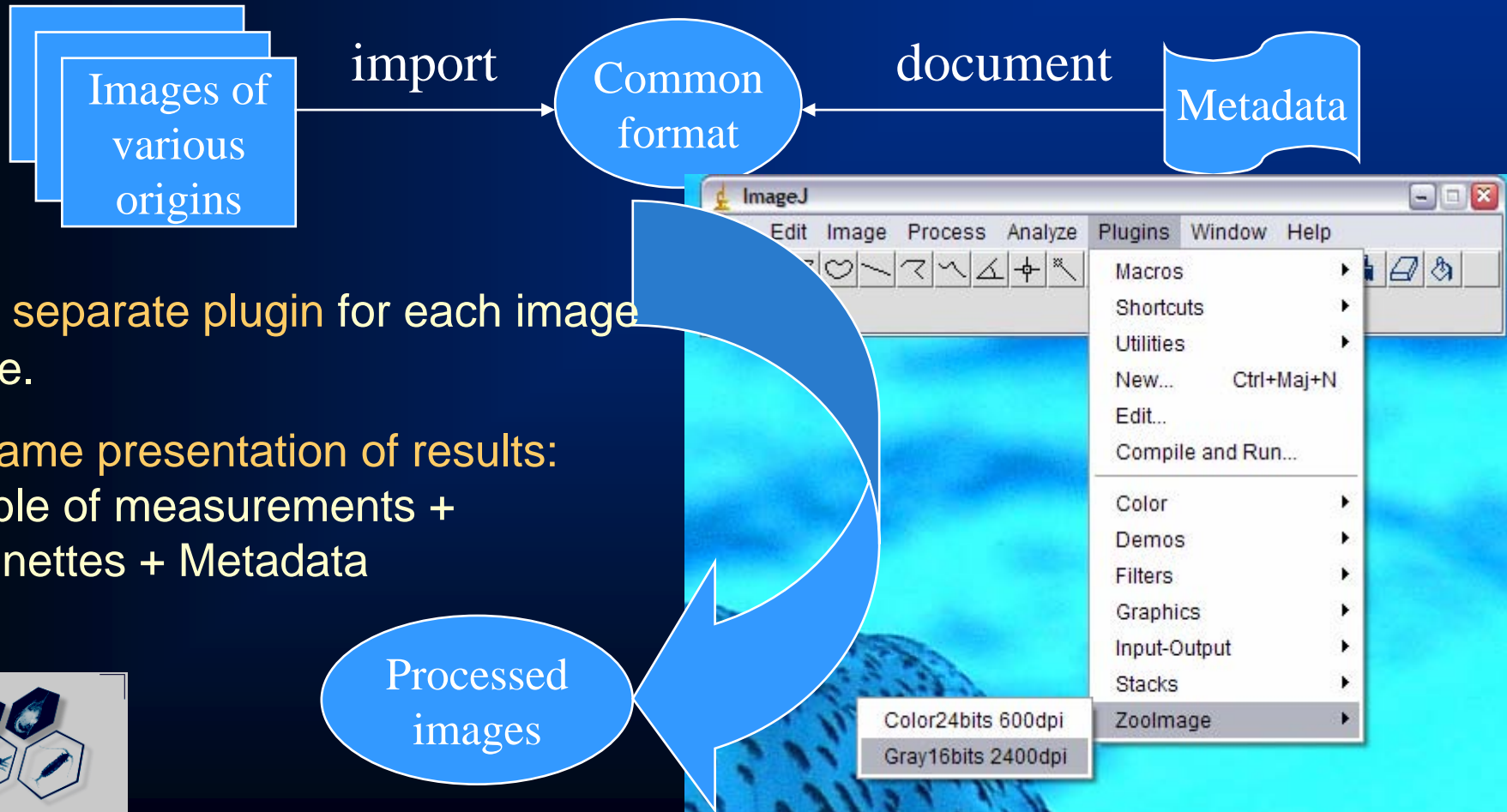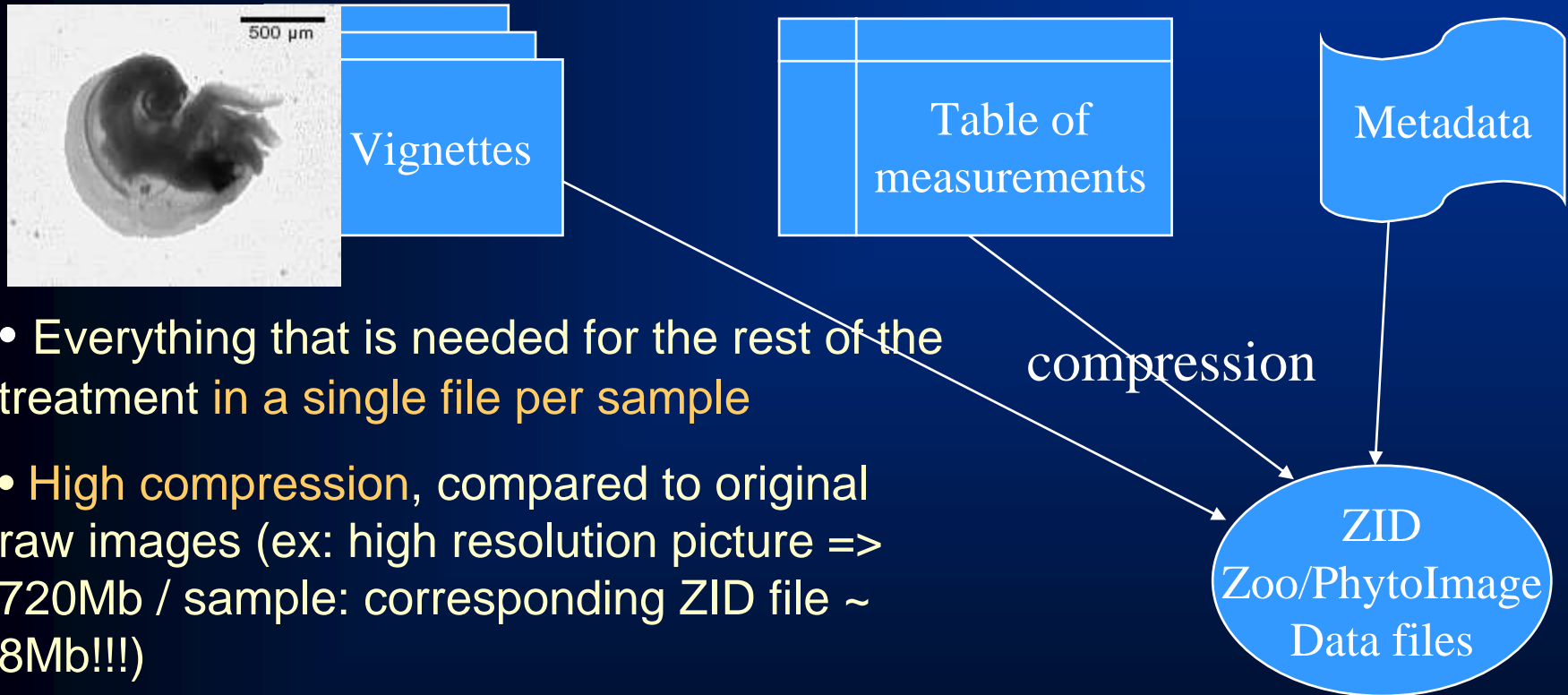
• A separate plugin for each image type.

• Same presentation of results:
Table of measurements +
Vignettes + Metadata

Processed images

ImageJ

Edit   Image   Process   Analyze   Plugins   Window   Help

Macros
Shortcuts
Utilities
New...   Ctrl+Maj+N
Edit...
Compile and Run...

Color
Demos
Filters
Graphics
Input-Output
Stacks
ZooImage

Color24bits 600dpi
Gray16bits 2400dpi

ZID files: common format & compression

Vignettes

Table of measurements

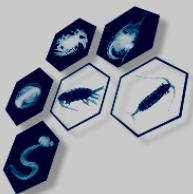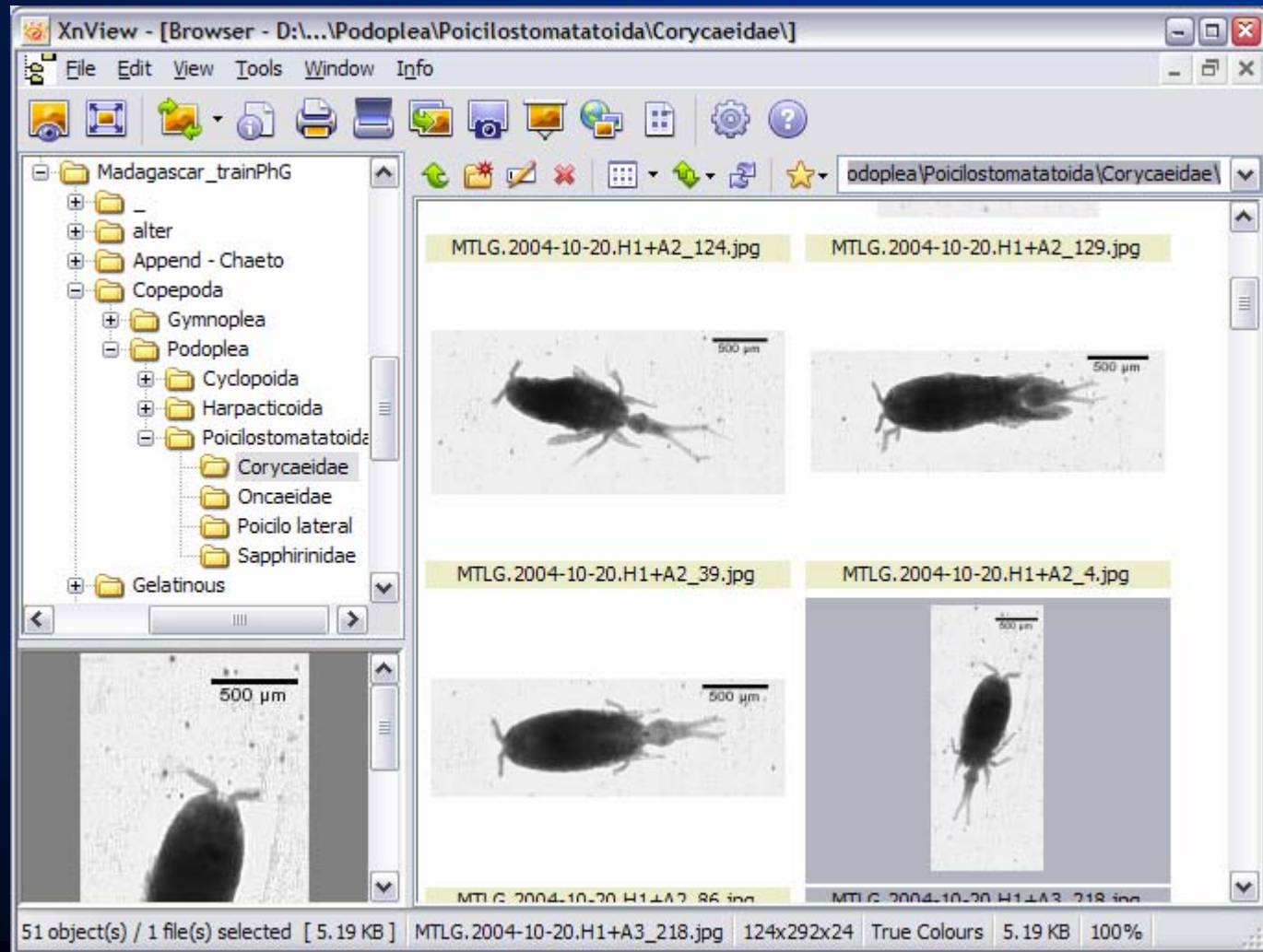**Metadata**

compression

ZID Zoo/PhytoImage Data files

• Everything that is needed for the rest of the treatment in a single file per sample

• High compression, compared to original raw images (ex: high resolution picture => 720Mb / sample: corresponding ZID file ~ 8Mb!!!)

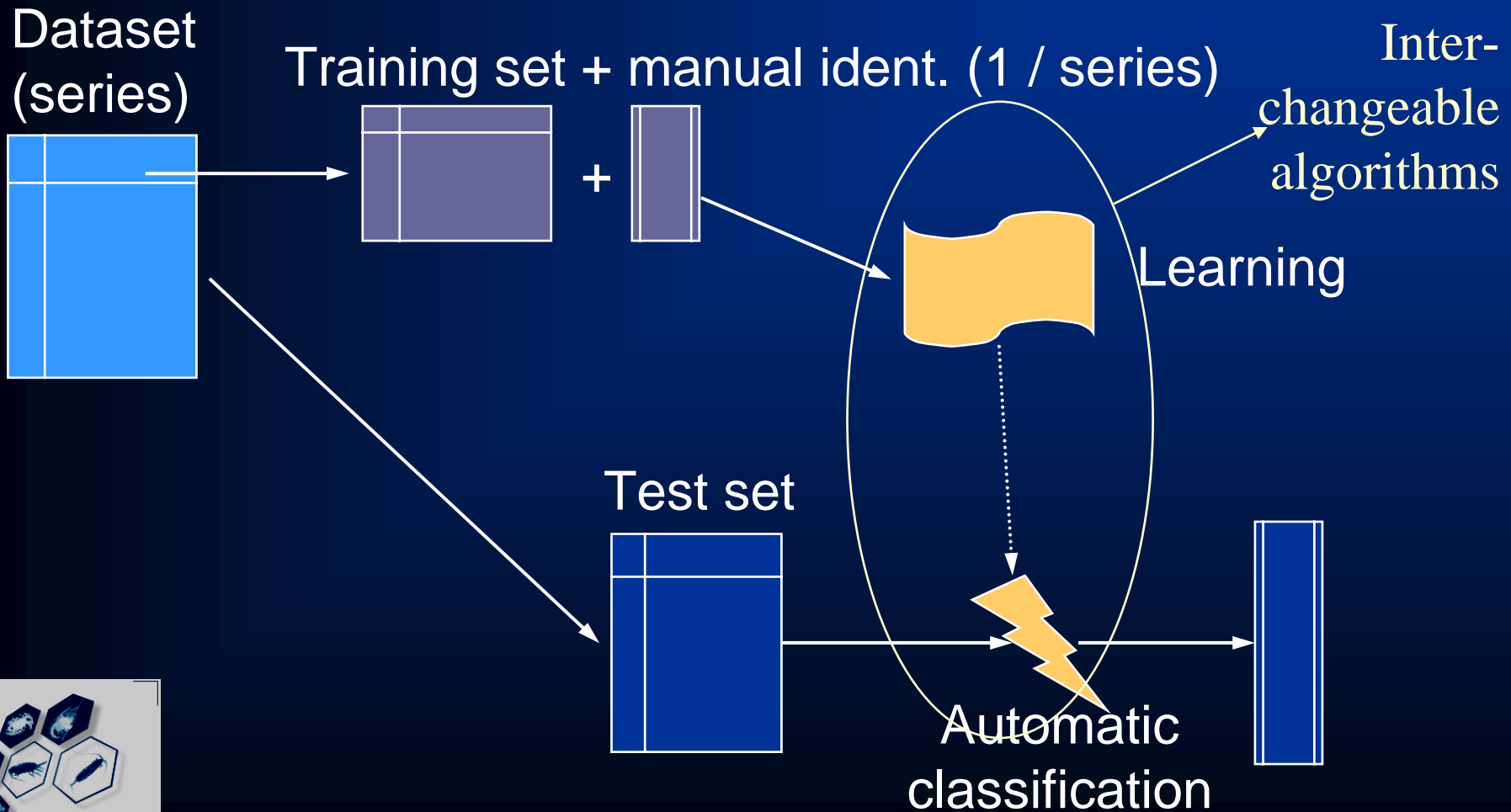*=> disk space problems for very large series / images are solved.*

Hierarchy grouping & manual training sets

• Any groups you want.

• Hierarchy reflect relationship – various possible levels of details.

• Taken into account automatically by the software.

Common framework for classifiers

Dataset (series)

Training set + manual ident. (1 / series)

+

Inter-changeable algorithms

Learning

Test set

Automatic classification

Computing of pertinent measurements

ZooImage computes (mixing of fractions, averaging over replicates and conversion per cubic meter of seawater are done automatically):

- Abundances per taxa (absolute, relative, log-scale),

- Biomasses, carbon content, etc.; total or per taxa (you must provide a conversion table),

- Size spectra; total or per taxa. Calculation of the slope of an exponential decreasing model. The number and width of classes are totally free (even irregularly spaced classes can be computed).
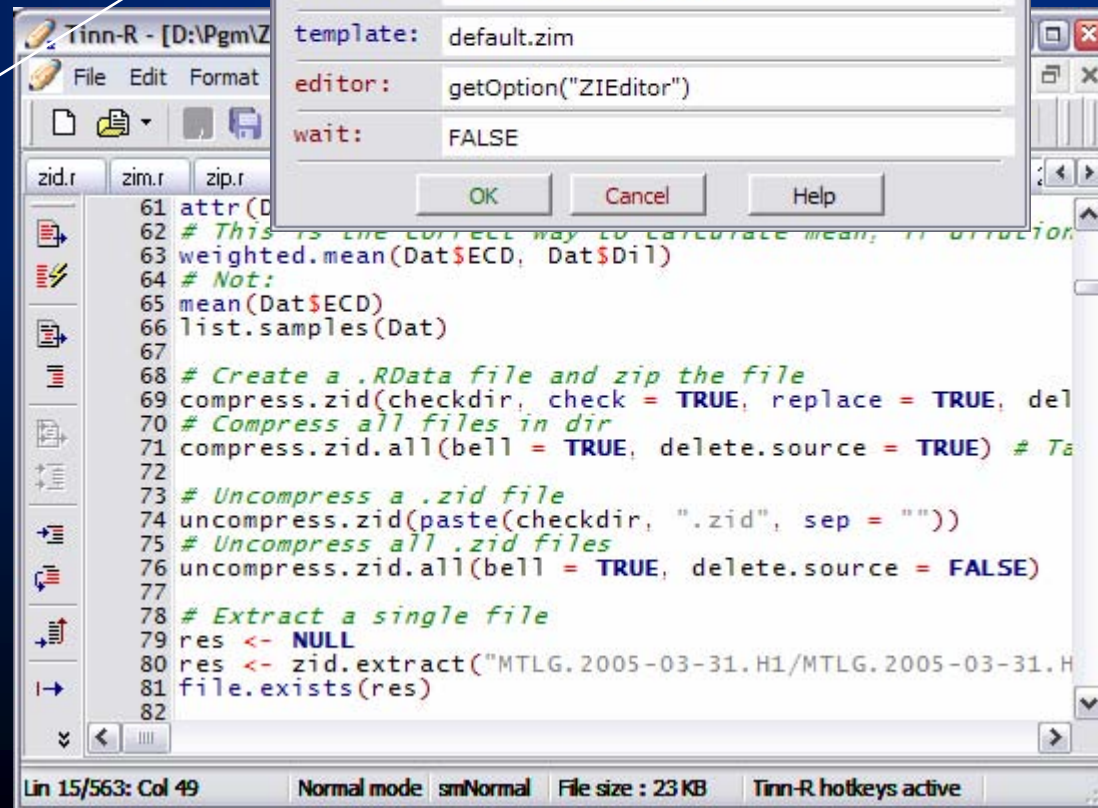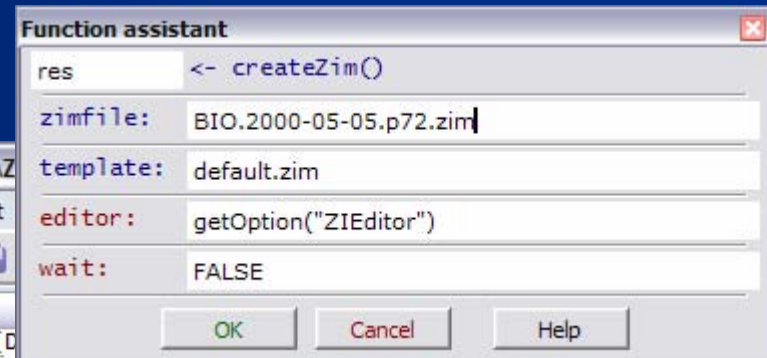
Accessibility

- Three levels, three user interfaces:

1. Full point&click assistant (complete beginner),

2. Dialog boxes prompting for command options (intermediate user),

3. Toolbox of functions for pure programmation (developers)

Current performances & future

- 20-30 taxa with 75-80% global accuracy on an average training set

- ZooImage analyses almost any kind of image, except Zooscan and VPR (for the moment)

- Developement highly dependent on contributions and actual use, but better algorithms, more graphs, more friendy GUI and a lot more documentation are planned