Emiliano Barbuto – Cristiano Bozza
Salerno Emulsion Group
LNF Oct. 2002

OPERA Scanning Cluster:
a working embryo of an IT infrastructure for emulsion scanning

Presentation Plan

⇒ Motivation

⇒ Overview

⇒ Batch Manager – orderly handling scanning tasks

⇒ A stand-alone scanning SW becomes a Scan Server

⇒ The heart of the system – a relational DB for scanning

⇒ Web Services I: managing scanning through the Web

⇒ Web Services II: remote data analysis

⇒ Next developments...

Motivation

The trigger to start this development was the Sept. 2002 test exposure, aimed at testing the Vertex Location in OPERA

Suddenly, we had to provide capability for unassisted long-running batches

The expected data size was much larger than before

Real-time data quality monitoring was needed

Technicians trained to CHORUS scanning had to face new scanning routine duties

Test .NET technology in OPERA-like everyday's work

Overview

Building blocks

Mic PC #1

Mic PC #2

Mic PC #3

Batch Manager

File Server

Web Interface

Local DB
(SQL Server,
ORACLE)

Scanning cluster operators only

Worldwide access through standard ways (HTML, ODBC)

Overview

All the Scanning Cluster software is being developed under .NET, mostly in C#
Code can be written in other .NET programming languages as well

The .NET Remoting Services eliminate the need to define communication protocols
Server programs expose their objects to network clients by a single instruction
Network client programs connect to server objects by a single instruction, and use
them as if they were locally instances of their classes

The .NET Remoting Services eliminate the need to define communication protocols

Web Services are all made with ASP.NET technology

The software is growing in an orderly manner

SySal.NET.beta2 is the code-name of the present version of the libraries

Batch Manager (BM) is a service application ("daemon" in UNIX language)

At startup the Batch Manager explores the scanning cluster configuration using Local DB info $\Rightarrow$ Centralized cluster administration through DB

• Scanning tasks are grouped in batches and sent to the BM.

• The BM registers batch progress into the DB
$\Rightarrow$ Orderly recover in case of HW failure

• The BM automatically starts post-DAQ tasks (systematic errors correction, fragment linking, etc...) and saves the data to the DB as soon as they're ready

All operations are logged $\Rightarrow$ Easy to discover malfunction causes

A stand-alone scanning SW becomes a Scan Server

SySal2000 architecture: COM+
Scanning Cluster architecture: .NET

Interfaced through 2 new plug-ins
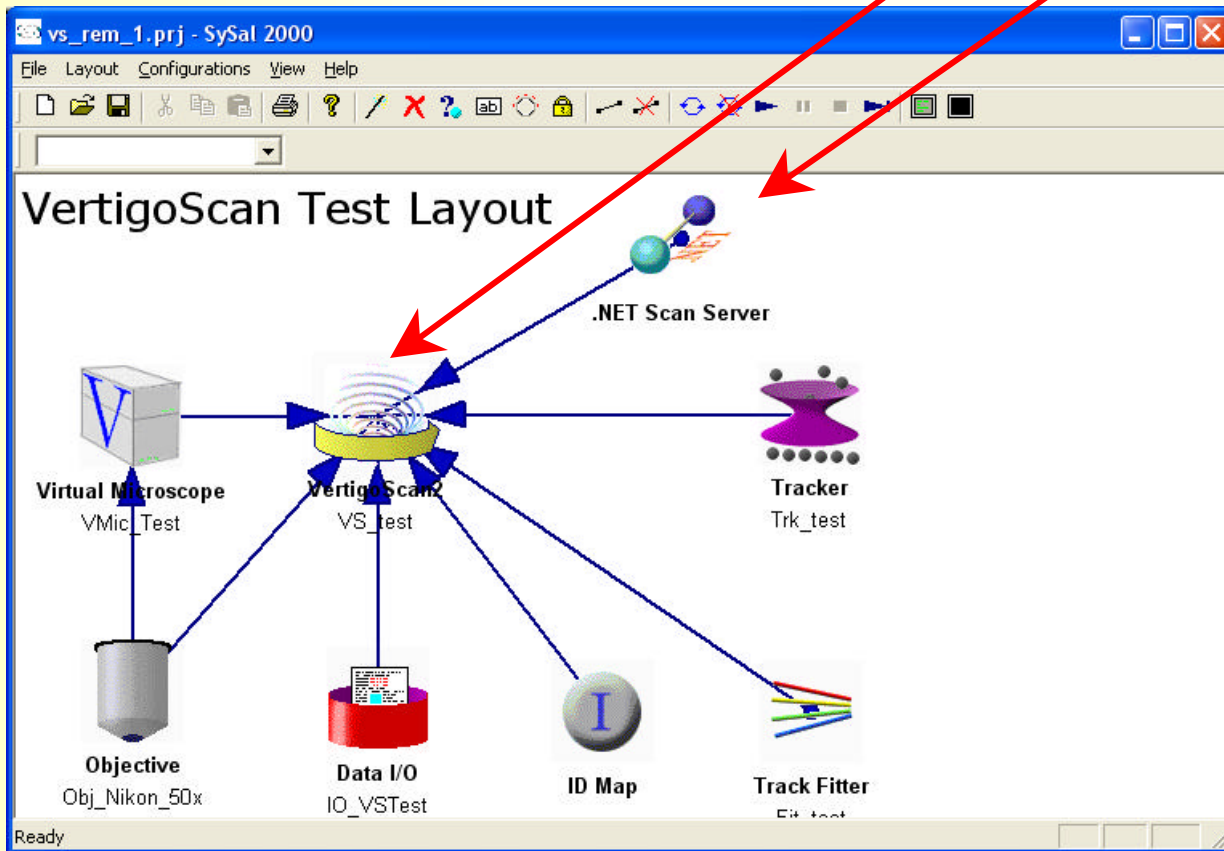
No additional configuration needed

Remote commands mimic human user operations

OPERA Local Scan DB alone contains the description of the cluster configuration, operating parameters and data altogether
$\Rightarrow$ Data quality and sources are fully documented in an unambiguous way

The DB is presently implemented by MS SQL Server 7.0 / 2000 using standard SQL only

ORACLE DB server already set up in Salerno, but can be used only for development because of license constraint, not for data processing...

However, porting should be quite easy

The whole Scanning Cluster can be switched from one DB to another by simply changing the DB connection parameters in the .config file of each application
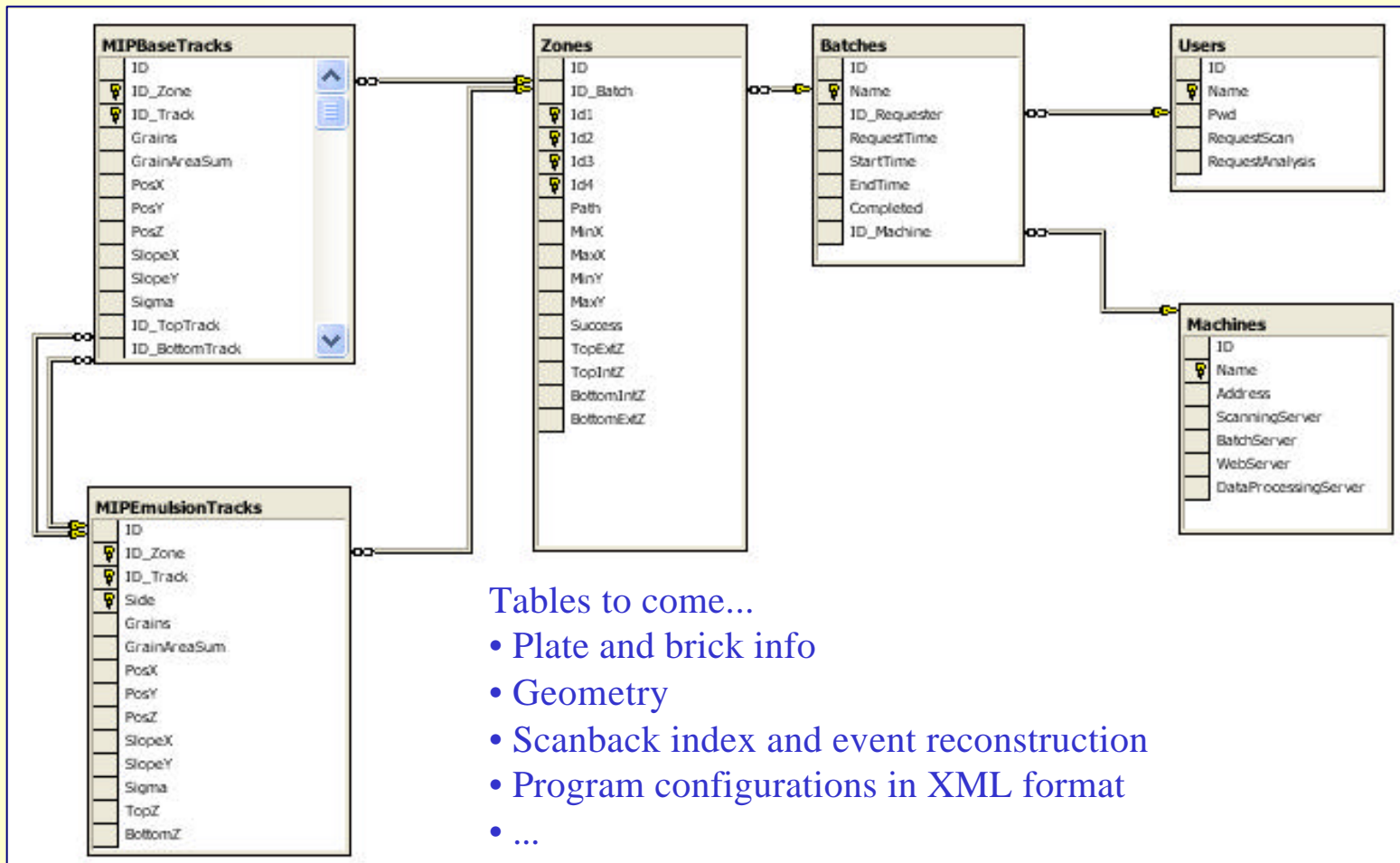$\Rightarrow$ No recompilation of software needed

OPERA Local Scan DB alone contains the description of the cluster configuration, operating parameters and data altogether

⇒ Data quality and sources are fully documented in an unambiguous way

MIPBaseTracks
- ID
- ID_Zone
- ID_Track
- Grains
- GrainAreaSum
- PosX
- PosY
- PosZ
- SlopeX
- SlopeY
- Sigma
- ID_TopTrack
- ID_BottomTrack

Zones
- ID
- ID_Batch
- Id1
- Id2
- Id3
- Id4
- Path
- MinX
- MaxX
- MinY
- MaxY
- Success
- TopExtZ
- TopIntZ
- BottomIntZ
- BottomExtZ

Batches
- ID
- Name
- ID_Requester
- RequestTime
- StartTime
- EndTime
- Completed
- ID_Machine

Users
- ID
- Name
- Pwd
- RequestScan
- RequestAnalysis

Machines
- ID
- Name
- Address
- ScanningServer
- BatchServer
- WebServer
- DataProcessingServer

MIPEmulsionTracks
- ID
- ID_Zone
- ID_Track
- Side
- Grains
- GrainAreaSum
- PosX
- PosY
- PosZ
- SlopeX
- SlopeY
- Sigma
- TopZ
- BottomZ

Tables to come...
- Plate and brick info
- Geometry
- Scanback index and event reconstruction
- Program configurations in XML format
- ...

The heart of the system – a relational DB for scanning

The data format for scanning output is very general

A successful test to convert UTS data and store them into our DB tables has been performed $\Rightarrow$ Useful to test algorithms and do cross-checks

Both ORACLE and MS SQL Server can be accessed worldwide through TCP/IP and ODBC $\Rightarrow$ Platform independent data availability

As all service applications working in background, the Batch Manager does not have a graphical interface

The Batch Manager is very easy to access programmatically, but in many cases operators want quick access to information

If a custom interface is developed, each computer that is being used to access the BM must be equipped with client software

However, PCs already come with a general purpose client software... The web browser!!!

Web Services I: managing scanning through the Web

The BM functions are exposed by a Web Server, not necessarily running on the same machine

Web Services II: remote data analysis

Remote Data Analysis was developed as a custom interface to interact with data recorded during scanning procedure, to monitor their quality and to perform a quick analysis.

Interacting with database...
MS SQL Server 7.0 - Oracle

Local DB
(SQL Server,
ORACLE)

Sql Instructions

RDA

Data Flow

Remote or
Local User

The local choice of the database
will have no consequences for the
"external user".

He will face no differences since standard html interface controls
(buttons, check and list boxes) will be used to code standard SQL
instructions to select data. In any case data will be presented in the
same way.
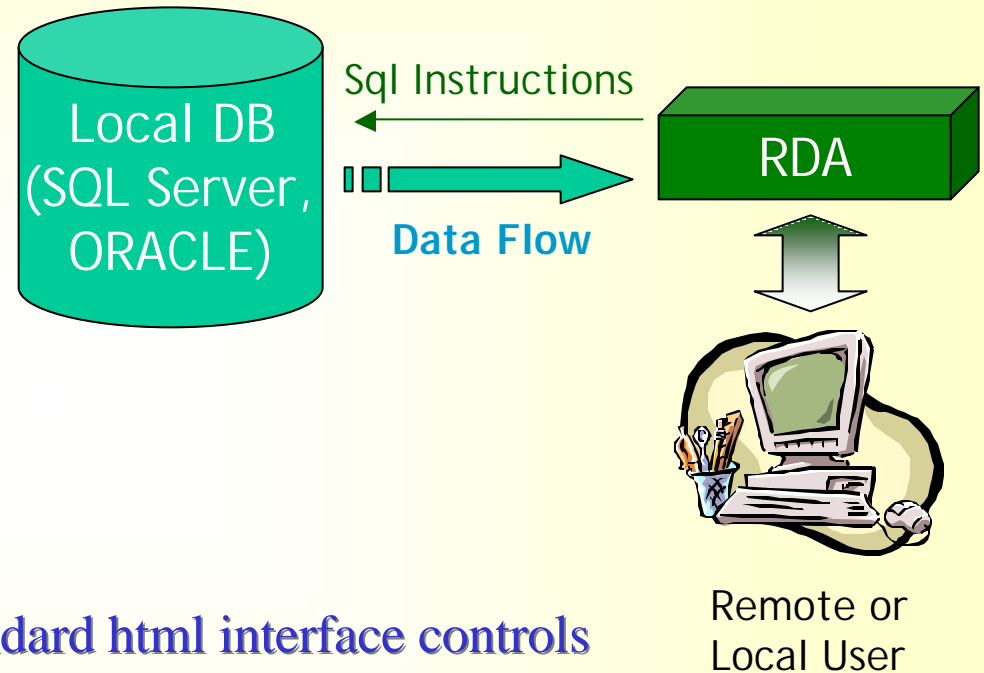
The next step towards user – database interaction.

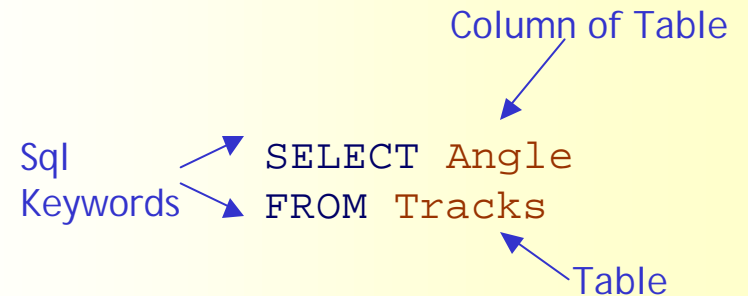Data selection using programming
languages like Fortran or C++.

SQL and other 4th generation languages.
Very easy and straightforward for database users.
Some graphical tools needed to display data.

Column of Table

Sql
Keywords

SELECT Angle
FROM Tracks

Table

Visual Interface.
A couple of "clicks" to select db contents, to
display them and to save plots for analysis.

Already implemented in Salerno
to select tracks for scanning
procedure in CHORUS
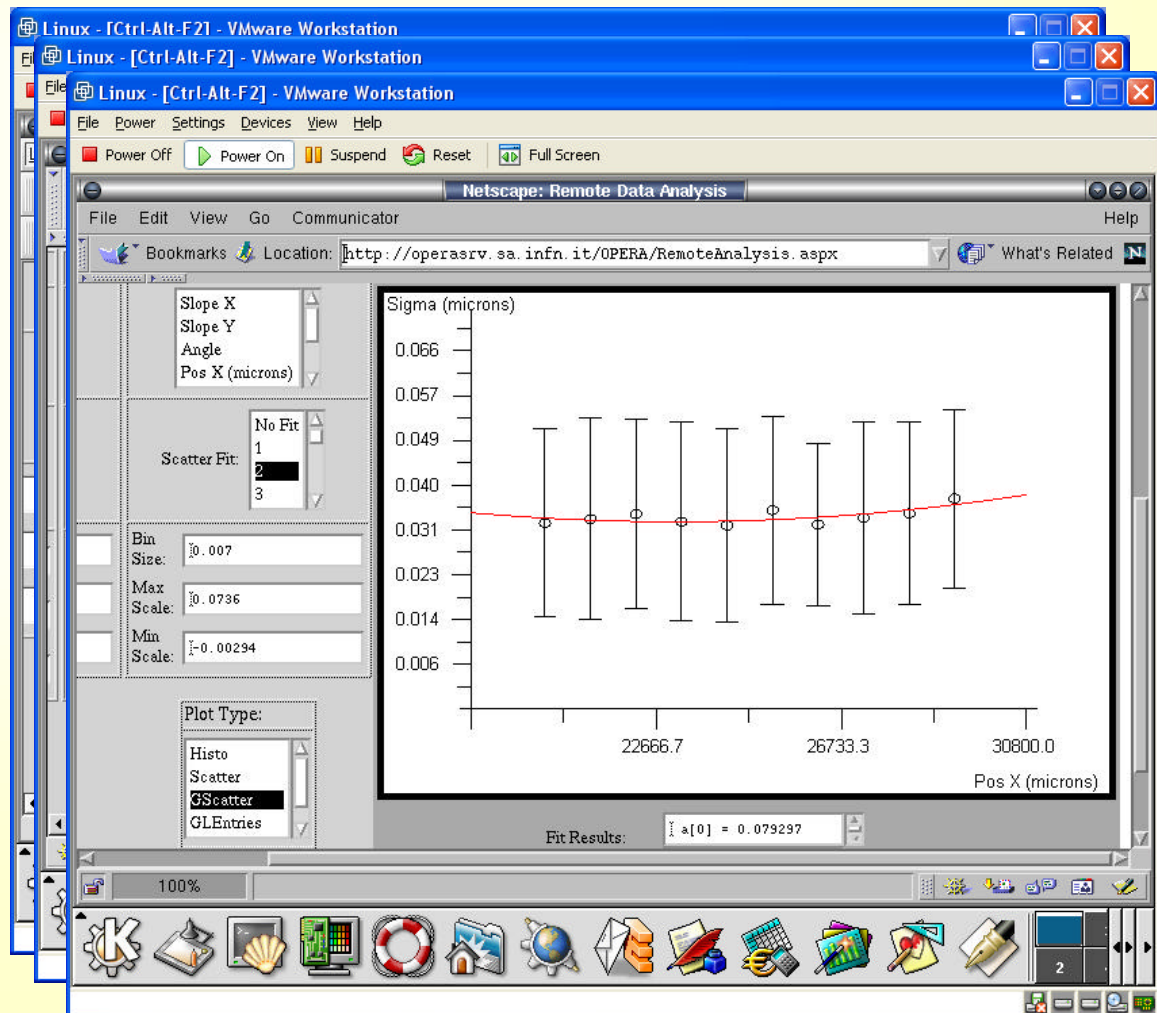
Web Services II: remote data analysis

Application compatibility was tested with several browsers.
The web page formatting and the user interface resulted to be only slightly influenced by the browser choice.

Microsoft Internet
Explorer (Windows);

Konqueror (Linux);

Mozilla (Linux);

Netscape (Linux);

Web Services II: remote data analysis

Interacting with scanning files...
(European Scanning System output files)

It can access all scanning files format:
• tlg format (thin emulsion sheet linked tracks);
• rwd format (raw data: single side emulsion tracks, track grains, corrections applied).

Every user can submit to the web service both remote scanning files (stored in the scanning cluster where the web service is installed) and local scanning files (stored on his/her own file server).
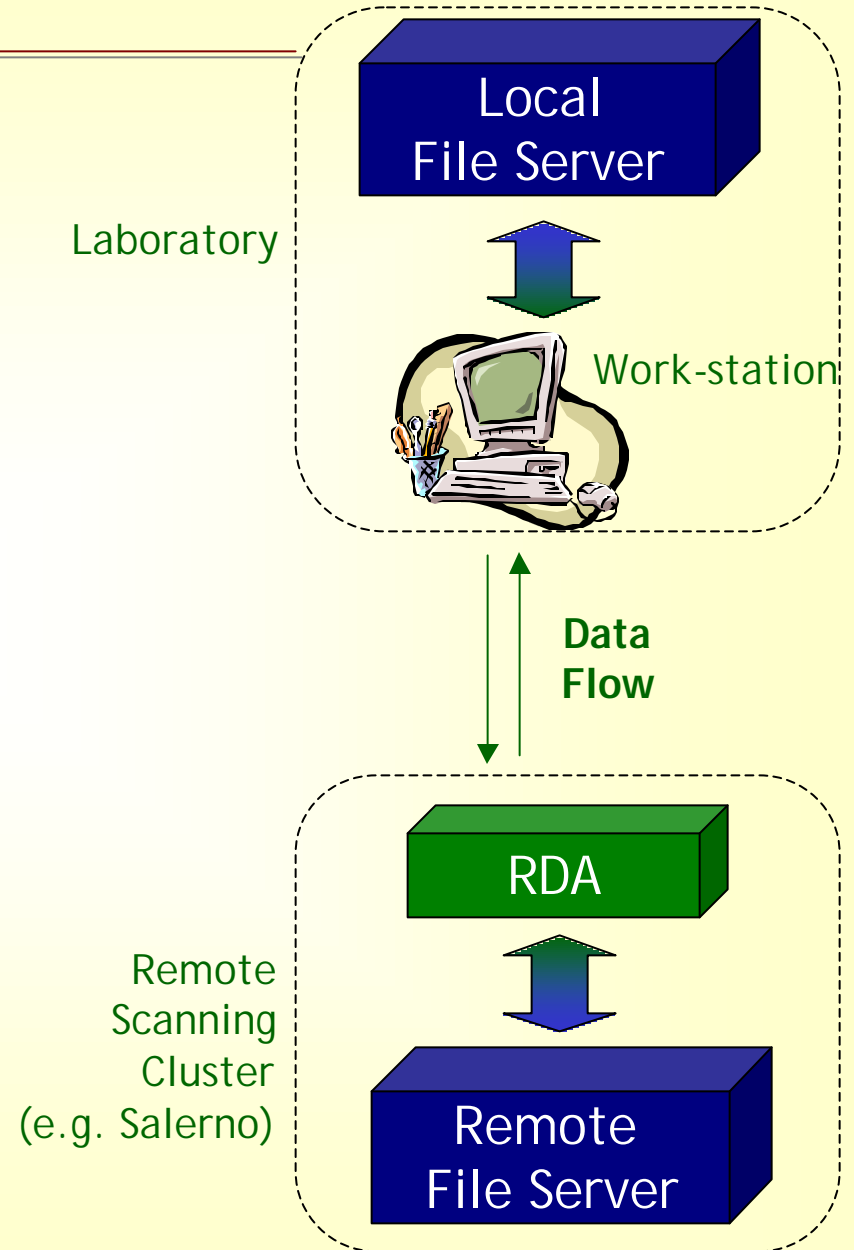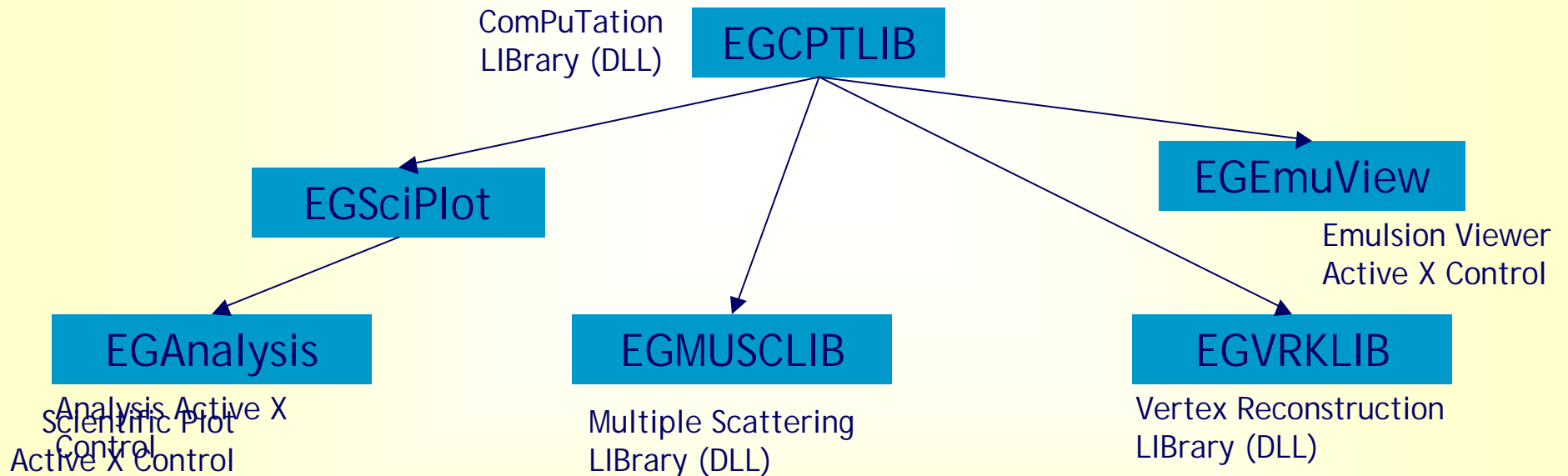
Technology behind this project...

Our emulsion group already developed some tools organized in a hyerarchical structure ("basic bricks", derived libraries etc...) using COM technology.

Web site: https://sysal2000.sa.infn.it/

ComPuTation
LIBrary (DLL)

EGCPTLIB

EGSciPlot

EGEmuView

Emulsion Viewer
Active X Control

EGAnalysis

EGMUSCLIB

EGVRKLIB

Analysis Active X
Scientific Plot
Control
Active X Control

Multiple Scattering
LIBrary (DLL)

Vertex Reconstruction
LIBrary (DLL)

Porting of this libraries into .NET framework started one month ago and now it is almost completed. They are written in C# and the structure was partially modified to exploit the .NET capabilities of developing stand-alone libraries and distributed applications.

Web Services II: remote data analysis

Computation Library

Library for data fitting and handling, for MC Generation, for on-line and off-line corrections.

Graphics Library

Library to plot histograms, scatterplots, 2d plots, 3d plots and data fitting.

Library output can be very general. It can plot in a specified area of a form (e.g. the display of an applicaton) or in a memory area (bitmap); this memory area can be then dumped in a file of a specified format.

```
Bitmap b = new Bitmap(Width, Height);

Graphics gPB = Graphics.FromImage(b);

gA = new GraphicsAnalysis.Plots();

gA.Histo(gPB, Width, Height);

b.Save(OutputFileName, ImageFormat.Gif);
```

Creates a bitmap (memory area) b having some width and height.

Sets Graphics library output to that bitmap.

Draws an histogram in the bitmap.

Saves the bitmap in an Output File image having gif format.

This is the fundamental idea of RDA. Plots are generated, according to graphics library, in some map areas, they are dumped into files and then shown into the browser window in one of Internet standard format (gif, jpeg or whatever).

Web Services II: remote data analysis

Graphics Control

If you choose a different destination for you graphics library you can create a form control.

A control is an object that performs some operations. You can choose to add this control to the application you are writing and use its performances. It will appear on the forms of your application as any other user interface.
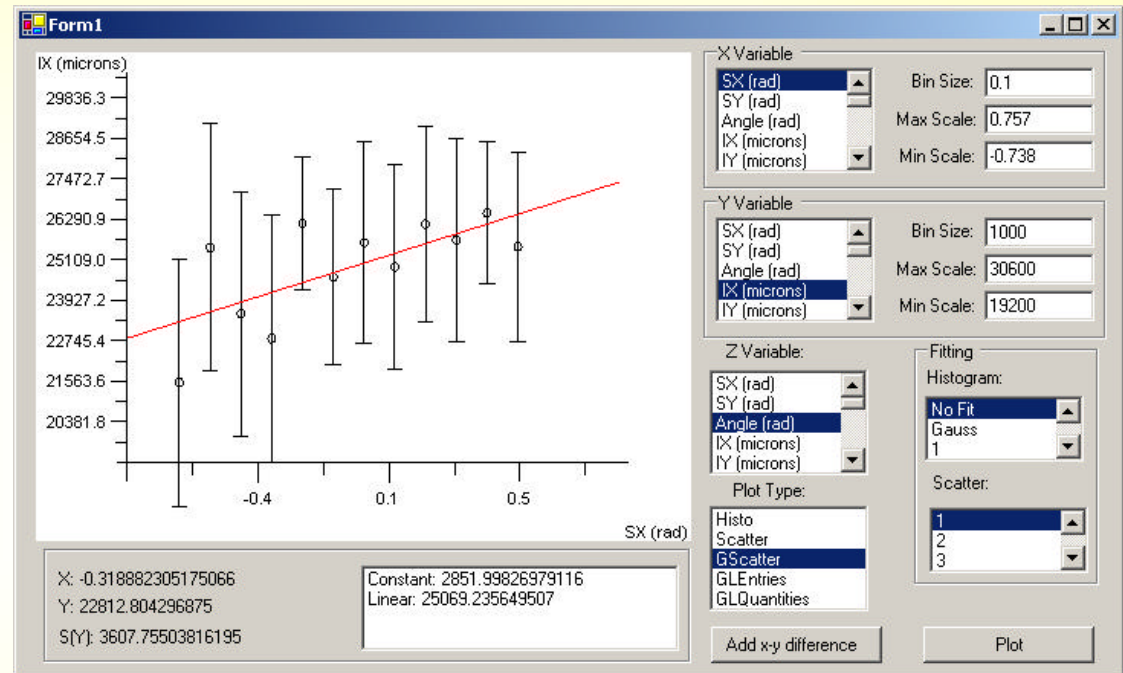

```
Graphics gPB = panel1.CreateGraphics();

gA.Histo(gPB, panel1.Width, panel1.Height);
```

The output of the graphics library is set equal to the white panel (in the code it is called panel1) of the form in the picture.

Web Services II: remote data analysis

Analysis Control

Data arrays (coming from files or database) can be loaded in the control which creates a matrix (n-tuple). A general analysis can be performed, data can be corrected and exported together with corrections applied.

These ideas were exported in a straightforward way into a web environment to create the RDA web service.

Multiple Scattering Library

This library is able to compute Multiple Coulomb Scattering to simulate it and to evaluate the alignment of emulsion sheets.

Next developments

The OPERA Scanning Cluster is still evolving

Its modular architecture allows us to upgrade it with time without stopping data taking and analysis

The DB is being ported to ORACLE and upgraded with many more tables

Data processing code and event reconstruction code are being ported into this new IT infrastructure

Hints for the Vertex Location logic and the decisional processes for Vertex Selection and Measurement, and event study will be among the products of this "embryonic structure"

Physics analysis can itself be performed through Web interfaces

Data processing code