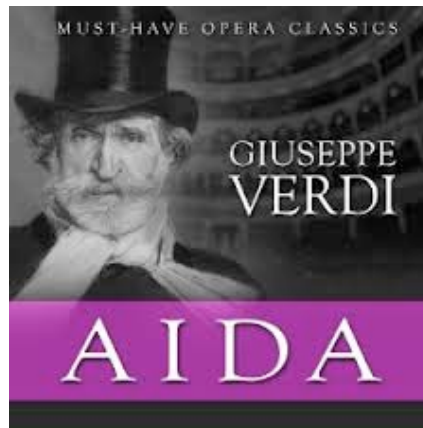


# AIDA



## Data analysis with R in an experimental physics environment

Andreas Pfeiffer (*CERN*) and Maria Grazia Pia (*INFN Genova*)

**IEEE NSS 2013**

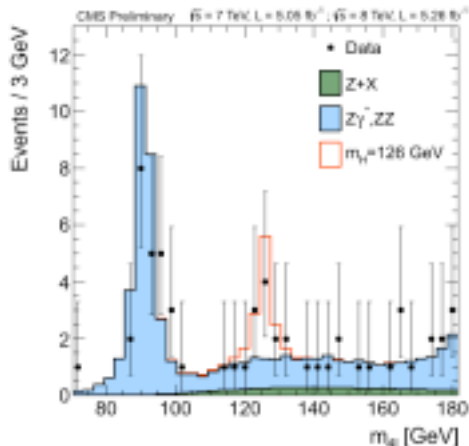
27 October – 2 November 2013

Seoul, Korea

**2013 IEEE NSS/MIC/RTSD**

*"Beyond Imagination of Future Science"*

# Daily work



© 2013 CERN, for the benefit of the CMS Collaboration (License: CC-BY-SA-3.0)

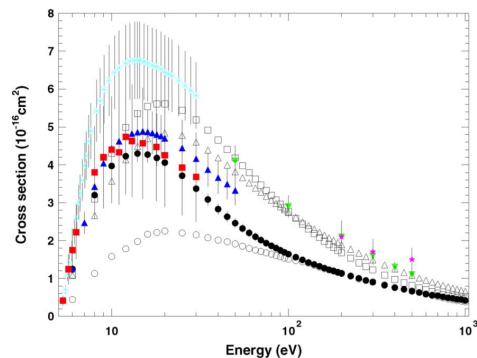
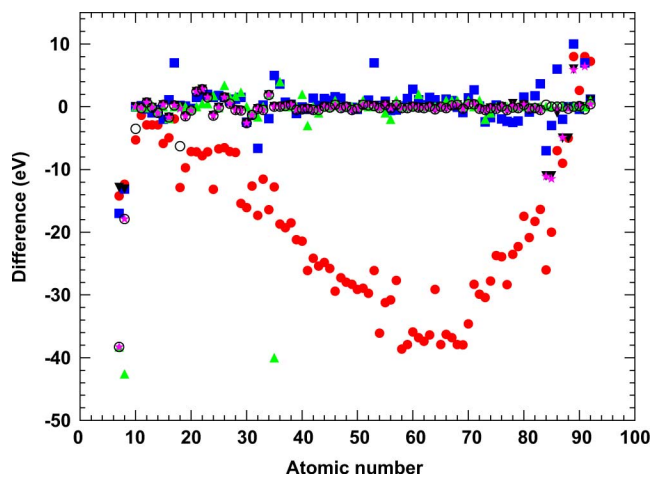


Fig. 24. Cross section,  $Z = 11$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [85] (green upside-down triangles), [86] (turquoise asterisks), [98] (pink stars), [99] (black circles), [100] (blue triangles) and [101] (red squares).

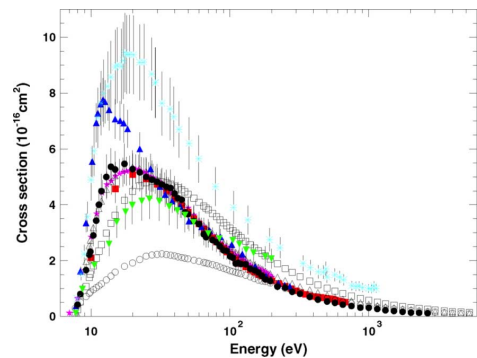


Fig. 25. Cross section,  $Z = 12$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [102] (pink stars), [103] (red squares), [104] (blue triangles), [105] (black circles), [106] (green upside-down triangles) and [107] (turquoise asterisks).

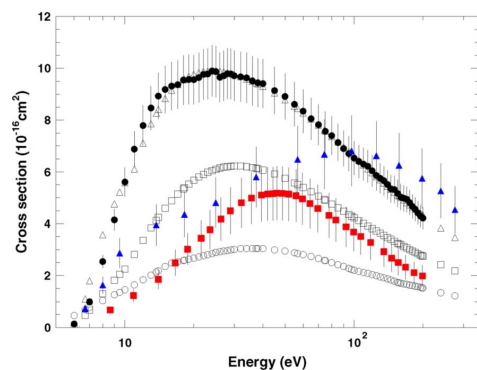


Fig. 26. Cross section,  $Z = 13$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [102] (black circles), [108] (red squares) and [109] (blue triangles).

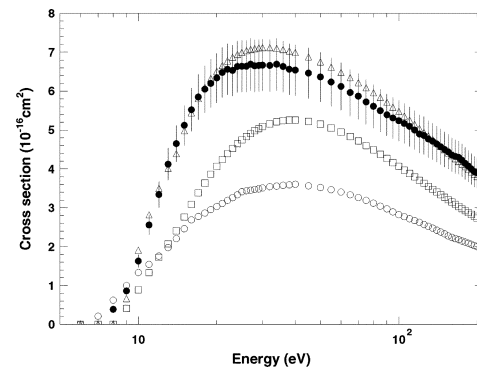


Fig. 27. Cross section,  $Z = 14$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [102] (black circles).

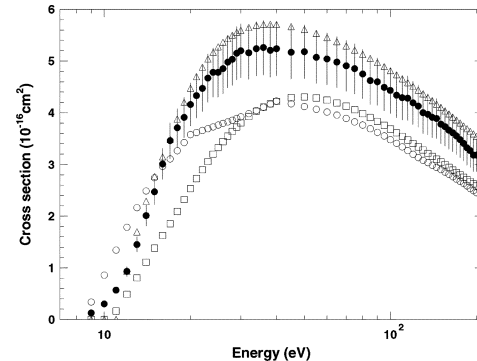


Fig. 28. Cross section,  $Z = 15$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [102] (black circles).

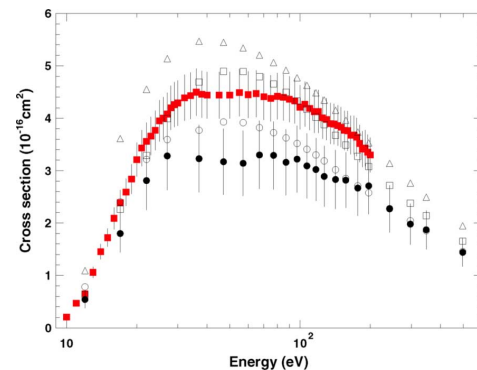


Fig. 29. Cross section,  $Z = 16$ : EEDL (empty circles), BEB model (empty squares), DM model (empty triangles) and experimental data from [102] (black circles) and [110] (red squares).

# Background

CERN Program Library Long Writeup W5013

CERN Program Library Long Writeups Y250

CERN Program Library Long Writeup Q121

In the old days...

*GEANT*

simulation

*HBOOK*

histograms, ntuples

*PAW*

analysis

Nowadays...

**Geant 4**

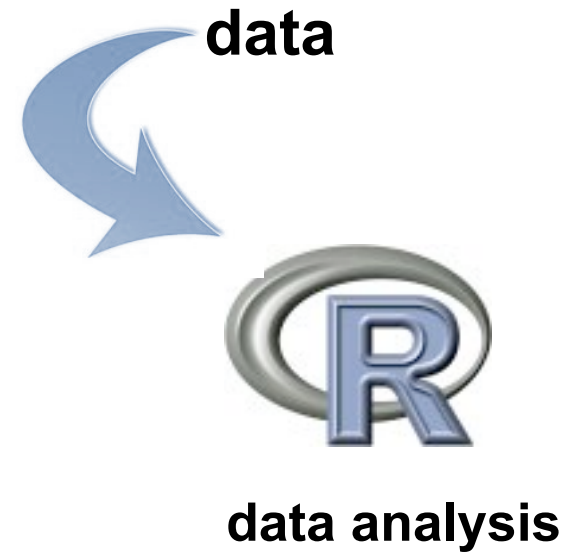
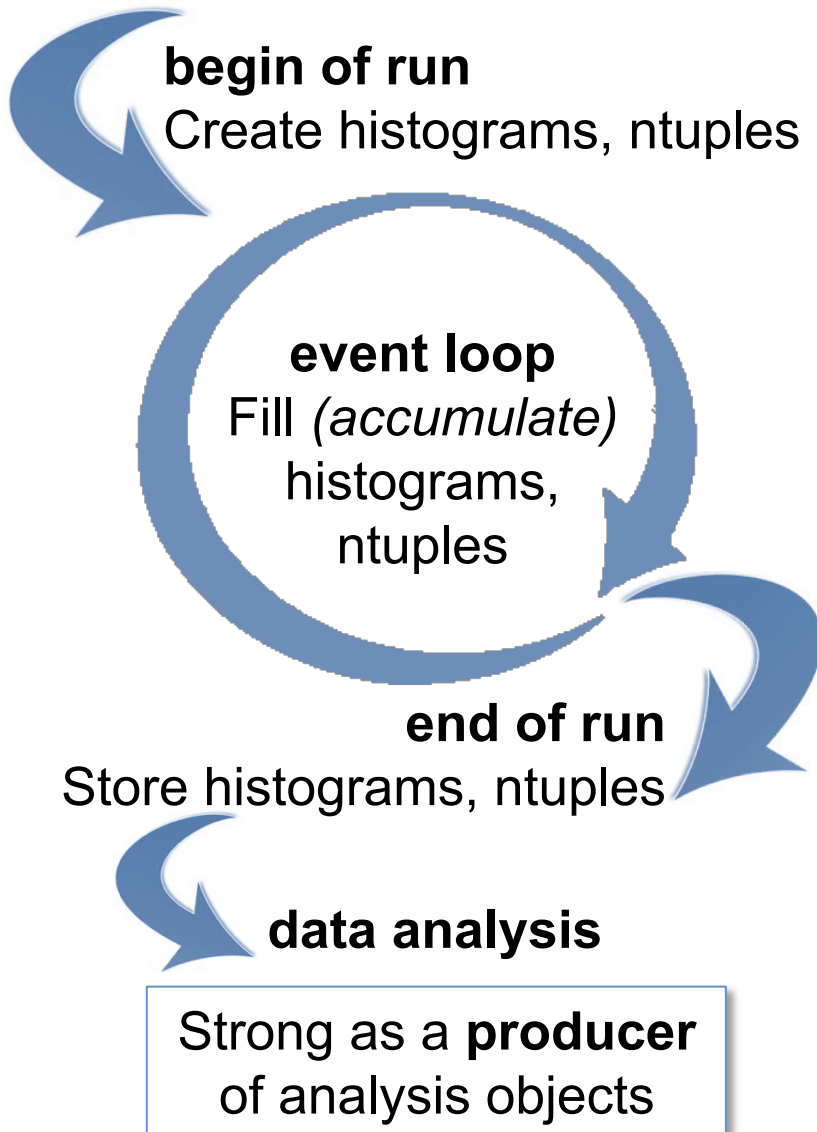
You are free to use  
whatever you want

Text file  
AIDA implementation  
ROOT  
...

You are free to use  
whatever you want

GnuPlot  
MATLAB  
iAIDA  
JAS  
Open Scientist  
PAIDA  
ROOT  
...

# Different conceptual models



Outstanding data  
**analysis capabilities**

# AIDA - Abstract Interfaces for Data Analysis

- Started in 1999, defining full set of interfaces
  - <http://aida.freehep.org>

“The goal of the AIDA project is to define **abstract interfaces** for common physics analysis objects, such as histograms, ntuples, fitters.

The adoption of these interfaces makes it easier for developers and users to select and use different tools without having to learn new interfaces or change their code.

In addition it is possible to exchange data (objects) between AIDA compliant applications through a **standard XML format.**”

# AIDA objects

## ● Data types

- Histograms 1D, 2D, 3D as statistical entities
  - also “dynamic”/”unbinned” ones (“Clouds”)
- Profile Histograms
- Ntuple
- DataPoints
  - vectors of “free form” N-dim data with errors

## ● Non-data types

- Annotations to add statistics/summary and “free form” info provided by user (key/value pairs)
- Fitter, Functions, Plotter, Analyzer

## ● Defined **XML format** for **data storage**

- “.aida” files (compressed XML)

# AIDA implementations

- Modular design to maximise flexibility
  - Factory pattern to create objects
  - Plugin modules for different implementations
    - e.g. native, Root, HBook stores to read/write histograms and tuples
- Implementations of interfaces in several languages
  - **C++**
    - **iAIDA** - <http://iaida.dynalias.net>
    - **OpenScientist** - <http://openscientist.lal.in2p3.fr>
  - **Java**
    - **JAS** (Java Analysis Studio) - <http://jas.freehep.org>
  - **Python**
    - **paida** - <http://paida.sourceforge.net>
- Flexibility through XML data interchange format
  - “.aida” files can be read by all across languages

# R

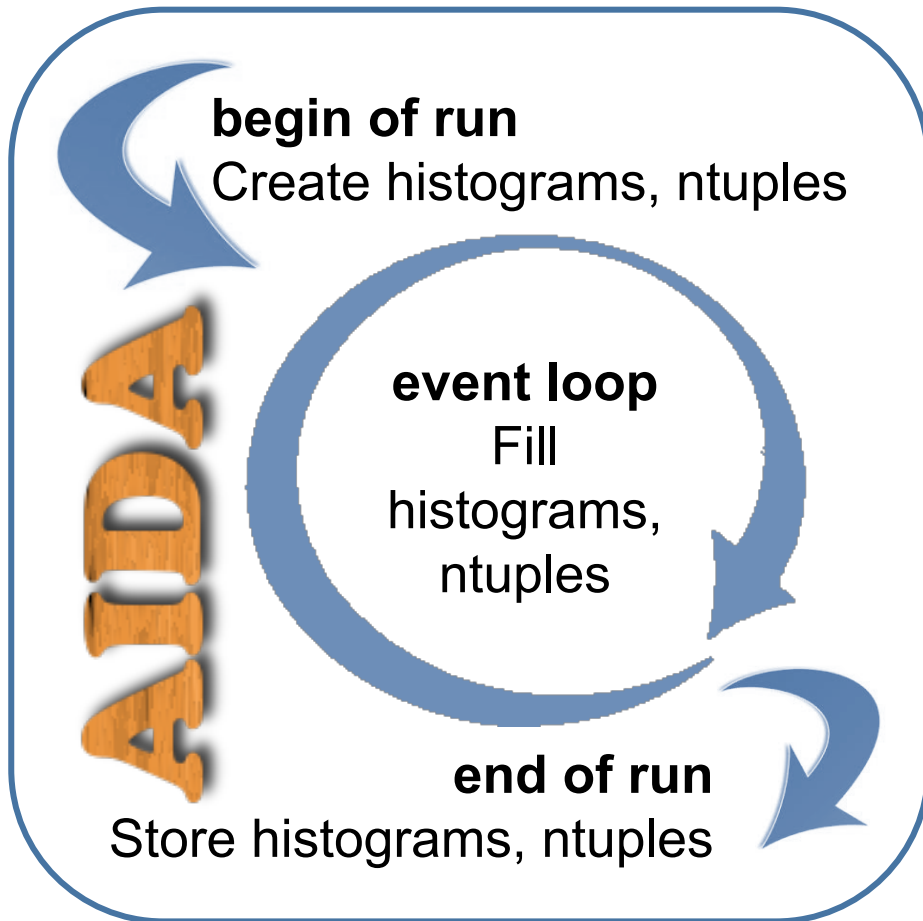
In 1998 John Chambers won the ACM Software Systems award for the S language, which the ACM heralded as having

**"forever altered how people analyze, visualize, and manipulate data"**

- R is a language and environment for statistical computing and graphics
  - Similar to S (can be considered as a different implementation of S)
- GNU project
- Provides a wide variety of **statistical** and **graphical tools**
- It is highly **extensible**
- Used by a **huge multi-disciplinary community**
- Strong at producing well-designed, publication-quality plots
- Runs on Windows, MacOS X, Linux (various distros)



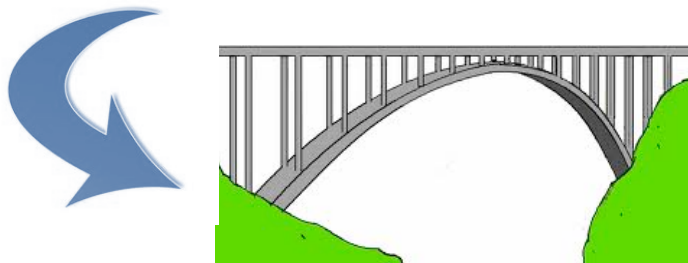
# Getting the best of both worlds



- A **lightweight** system for dealing with analysis objects in experimental software scenarios

- A **powerful**, extensible data analysis system

- A **transparent** bridge between the two



**data analysis**

# aidar - Interfacing AIDA with R

- Interface to read AIDA XML files into R
  - Exploiting the power of R for analysis
  - Using the existing XML package in R
- **aidar** converts AIDA objects from (XML) file into **data.frames**
  - Histograms, Clouds, Profiles, Ntuples
  - *getFileInfo(<fileName>)* to get overview
- Developer version available from github:
  - <https://github.com/apfeiffer1/aidar>
- Easy install via *devtools* package (see *Readme on github*)
- Plan to have it as regular **CRAN module** by end November

Seamless data production and analysis, transparent use in R

# Initialization (start of run) AIDA

e.g. **iAIDA**

```
// Create the analysis factory and the tree factory
```

```
...
```

```
// Creating a tree mapped to a new XML file
```

```
std::auto_ptr< AIDA::ITree > tree( tf->create( "comptoncs.xml", "xml", readOnly, createFile, "uncompressed" ) );
```

```
// Create a tuple factor and a histogram factory
```

```
...
```

```
// Create a 1D histogram
```

```
AIDA::IHistogram1D* hSigma = hf->createHistogram1D("10","Cross section", 100,0.,1.);
```

```
// Create a ntuple
```

```
AIDA::ITuple* ntuple = tpf->create( "1", "Compton cross section", "float z, e, lib, pen, std" );
```

## Event loop

```
// Do your calculations in the event/track loop
```

```
...
```

```
// Fill histogram
```

```
hSigma->fill(sigmaEPDL);
```

```
// Add data row to Ntuple
```

```
ntuple->fill( ntuple1->findColumn( "z" ), z );
```

```
ntuple->fill( ntuple1->findColumn( "e" ), e );
```

```
ntuple->fill( ntuple1->findColumn( "lib" ), sigmaEPDL );
```

```
ntuple->fill( ntuple1->findColumn( "pen" ), sigmaPenelope );
```

```
ntuple->fill( ntuple1->findColumn( "std" ), sigmaStandard
```

```
ntuple->addRow();
```

## End of run

```
// Committing the transaction with the tree
```

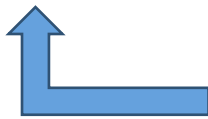
```
tree->commit();
```

```
tree->close();
```

# R session

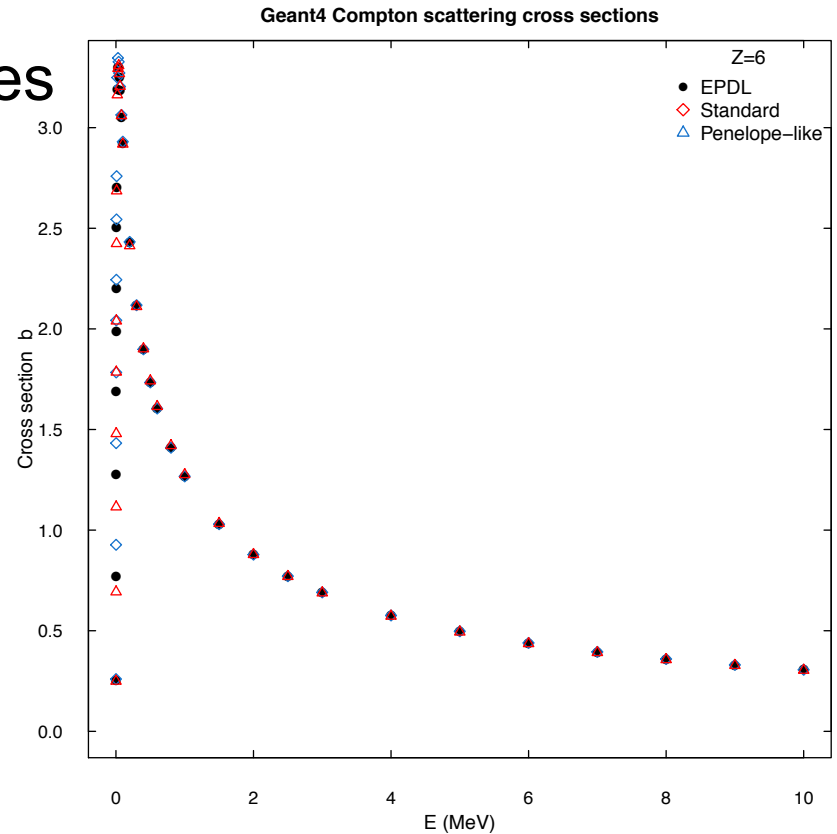
Load devtools and aidar packages

- `histoFile = "comptoncs.xml"`
- `t1 = getTuple(histoFile, '1')`



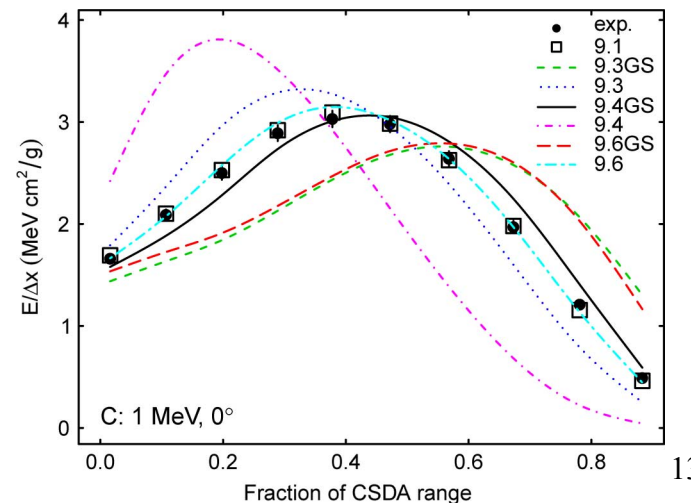
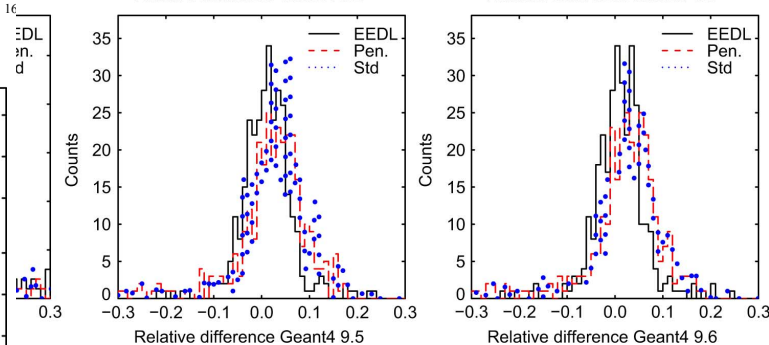
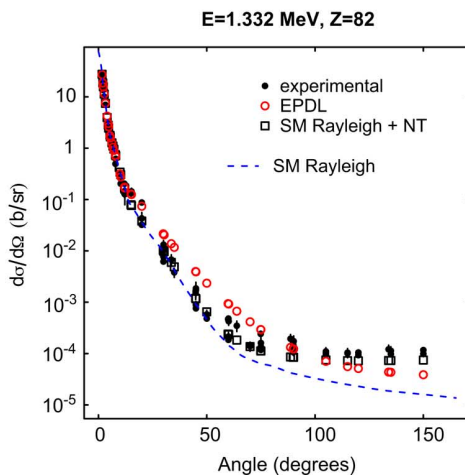
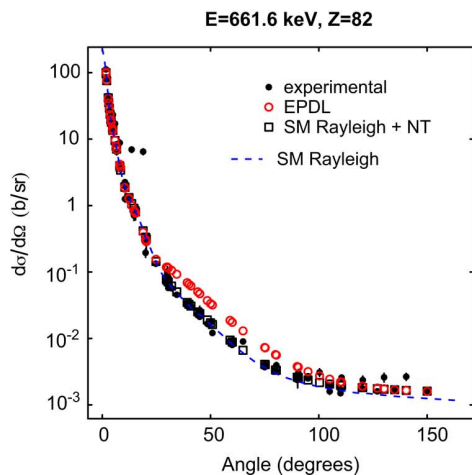
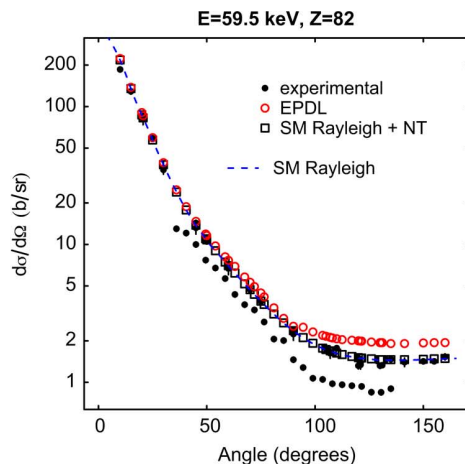
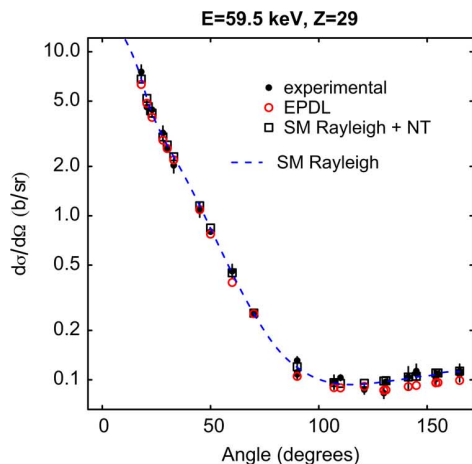
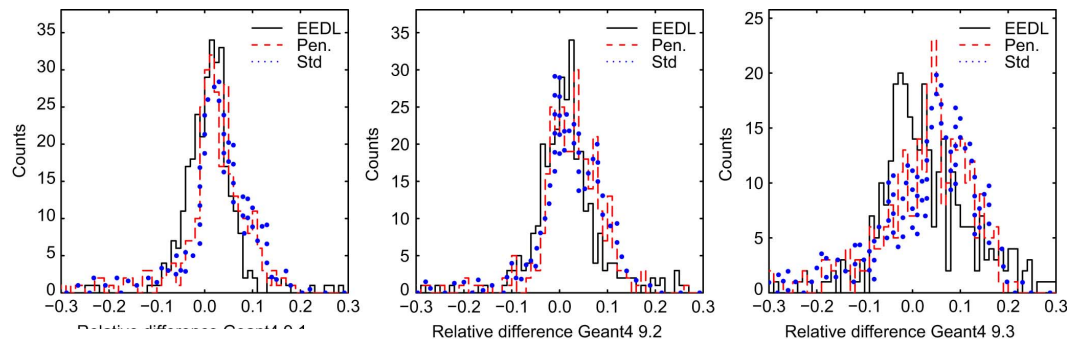
*T1 (AIDA ntuple)  
gets converted into  
a R data.frame*

- `plot(t1$e, t1$lib, ...)`



# Recent Geant4 validation

BATIĆ *et al.*: PHOTON ELASTIC SCATTERING SIMULATION: VALIDATION AND IMPROVEMENTS TO GEANT4



# It works!

1636

IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 59, NO. 4, AUGUST 2012

## Photon Elastic Scattering Simulation: Validation and Improvements to Geant4

Matej Batič, Gabriela Hoff, Maria Grazia Pia, and Paolo Saracco

2934

IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 60, NO. 4, AUGUST 2013

## Validation of Geant4 Simulation of Electron Energy Deposition

Matej Batič, Gabriela Hoff, Maria Grazia Pia, Paolo Saracco, and Georg Weidenspointner

### This conference:

- **N29-4**, Physics Methods for the Simulation of Photoionization
- **N29-5**, Validation of Compton Scattering Monte Carlo Simulation Models
- **NPO2-141**, Validation of Geant4 Electron Pair Production by Photons

# Conclusions

- Bridge between two conceptually different data analysis scenarios
  - Addresses typical use cases in daily experimental practice
  - Best of two worlds
- Transparent to users

**AIDA**  
**iAIDA**

Lightweight, robust analysis system for data production



Powerful system for data elaboration and graphics

**Use it!**