CANSAS

WHY????

Ron Ghosh ILL
Pete Timmins ILL
Roland May ILL
Claudio Ferrero ESRF
Wim Bras
ESRF/DUBBLE

Outcome CANSAS meeting February 1998

- ~ 30 attendees from major central facilities (US, Europe, Japan)
- Inventory of existing data-analysis packages
  Richard Heenan, r.k.heenan@rl.ac.uk
- Inventory of data formats
  Malcolm Capel, capel@bnl.gov
- Attempt of collecting software to translate
  Dieter Schneider, schneider@bio.bnl.gov
- Inventory of what should go in datafile
  Adrian Rennie, rennie@gordon.cryst.bbk.ac.uk

- short satellite meeting on SAS XI Brookhaven 1999

- website: http://www.ill.fr/lss/canSAS/main.html

What do we see as the problem?

- several user groups use more than one facility and would like to use the same analysis software for all their data sets

- several user groups rely on 3rd party software (Glatter, Svergun etc.) to analyse their data

- Exchange of data sets between software packages is not always straight forward, especially not if a group doesn't have a computer wizard available

⇒ How do we get data from one software packet to another?

Additional problem:

There is no real inventory of what software packages people are using to analyse their data.

Now what are the fields that we're talking about:

- Conventional SAS producing 1D data files
  (relatively small files)

- Time-resolved SAS producing series of 1D data files
  (reasonable sized)

- Conventional fibre diffraction producing 2D data files
  (reasonable sized)

- Time-resolved fibre diffraction producing series of 2D data files
  (huge amounts of data)

Often the output of analysed 2D data is further processed using 1D analysis software

Who are our 'customers'?

Users of large facility SAXS and SANS instruments

Daresbury
200 days beamtime/year
3 SAXS stations
Average stay of user 2-3 days
Average number of visits/year 2-3

$\Rightarrow$ 80 user groups

Similar for Hamburg, Photon Factory, ILL
Slightly less for ISIS, Brookhaven, Riso
Unknown for several other places
(Hahn-Meitner, Riso, Lure, Elettra, Oak Ridge etc.)
When APS, ESRF and Spring 8 will be fully
operational 100 groups/machine more

$\Rightarrow$ 800 -1000 user groups

steps

1 - production of data
very instrument dependent how the data comes out
somewhere a data format is defined often defined by the local hardware/software and prejudices of the programmer

2 - reduction (background subtraction, normalisation etc.)
software requires an input format and either outputs data with same format or defines new format

3 - analysis
software again requires an input format and produces data with same format or defines new format

4 - comparison with modelled data
software again requires an input format

At each central facility the step between 1 and 2 will be taken care of.

But between 2, 3 and 4 not necessarily.

Now the problem can be solved by everyone using a standard data format but even the experts in this field are slightly divided. So one day this may become a reality but in the meantime it is a dream and we still face the problems described here.

What we would like to achieve:

- Enhanced possibilities to swap data around

How would we like to achieve that:

- Unified data format (long term option and this meeting is not called to define things on this)

- Data translation modules (short term option for people who are working now and want results now and not tomorrow)

- Publish the data formats in a central place and convince authors of packages that they want to be available to the public that they should stick to these data formats

- Convince authors that they should write different input routines and output routines that at least can handle some different used formats