

Access to Datasets

V.Ivanchenko
17th Geant4 Workshop
10-14 September 2012
Chartres, France

CMS request to reduce size of data files

- Studying migration of CMS software to Geant4 9.5 a problem have been identified:
 - HP data structure size was increased by factor 5
 - This make a problem for software distribution
- I make a survey of existing data structures:
 - G4EMLOW2.27 – 72.6 M
 - G4RadioactiveDecay3.4 – 6.0 M
 - G4PhotonEvaporation2.2 – 30.9 M
 - G4NEUTRONXS1.1 – 4.2 M
 - G4NDL3.14 – 223.7 M
 - G4NDL3.15 – 1064.7 M
 - G4DNL3.16 – 803.9 M
 - G4NDL4.0.PUB – 1432.5 M
 - G4NDL4.0.PUB (all files compressed) – 411.9 M
- Can we consider situation as an issue and think about reading of compressed files?

What is done for CMS development release

- Currently for CMS thermal data are removed
 - Size reduced from 1.4 GB to 1.0 GB
 - Method is not appropriate for CMS software, because data are taken not from official Geant4 site
- Tatsumi have studied other variants toof size reduction and all not so effective and are more complicate

Proposal

- Thermal data may be separated out of G4NDL into a specific data set
- We should try to read zipped files
 - We will reduce the size by factor 2-3
 - Not a problem to read file.gz using zip library already used by Geant4
 - Is it possible not to compress individually each file but to compress directory structure?