Electronic Data + Scanning Feedback Data Base

Roumen TSENOV,

Chorus Collaboration meeting

CERN, 7-8.12.1999

Demand from physics analysis:

“We need a tool for fast simultaneous access and manipulation of electronic data and scanning data for the same event!”

Inputs

Electronic data

- **ZEBRA structure (FZ files)**
  - **- raw events (DAQ)**
  - **- processed events**
    - **- CHORAL - few passes**
      - **- maxi-DST**
        (~ 1 TB for 94/97 data)
      - **- mini-DST (STEV bank)**
        (~ 10 GB for 94/97 data)
    - **- CHANT**
      - **- maxi-DST**
      - **- mini-DST (to be defined!)**
- **files are on tapes**

Scanning data

- **Scan-back procedure data (finding track on CS, SS and in bulk emulsion)**
- **Scanning one track (per event) in bulk emulsion (Phase I)**
  - Every scanning lab has its own format
- **Net-Scan data (Nagoya and CERN)**
  - data are stored in Objectivity/DB™
  - similar formats in both Labs
  - vertex is located in ~ 10% of events and they are NetScan-ed).
  - data volume: ~1.4 Mb/event raw data from which ~300 kb/event are left after some preliminary track finding.

Requirements

- **Simultaneous access to data both from electronic detectors and scanning in one application;**
- **Fast access to (*a selected subset of*) data;**
- **Access to data through FORTRAN code, giving access to a ZEBRA structure in memory;**
- **Possibility for accessing the data through C++ code as well** *(applications based on object oriented approach, natural access to NetScan data,…).*

Software infrastructure being developed to answer the demand

- **Chorus scanning feedback data base**

  - commercial software product* - Objectivity/DB™ *- has been chosen

  supports storing of very large volumes of data;

  different parts of the federated data base can be stored on different machines running different operating systems

  good support at CERN (RD45, basic solution for LHC experiments)

  access through C++ and Java applications

  Nagoya and CERN are using it successfully

- **300 Gb of disk space**

  - asked to keep there a subset of (scanning + processed electronic) data

- **Tools for storing/accessing data**

Data volume to be kept on disk

Input: 2×10^6 emulsion trigger events

- **NetScan data : ~ 10 % of all events, 300 kb/event → 60 Gb**
- **All CHORAL mini-DST: 10 Gb (I.M.)**
- **All EDIR files: 1 Gb (I.M.)**
- **CHANT mini-DST with re-processed data: ~100 Gb**

(10 times CHORAL mini-DST)

We are safely below 300 Gb.

Note: Chorus Oracle data bases for Phase I scanning data and CHORAL/CHANT processing bookkeeping (chinfo) reside elsewhere.

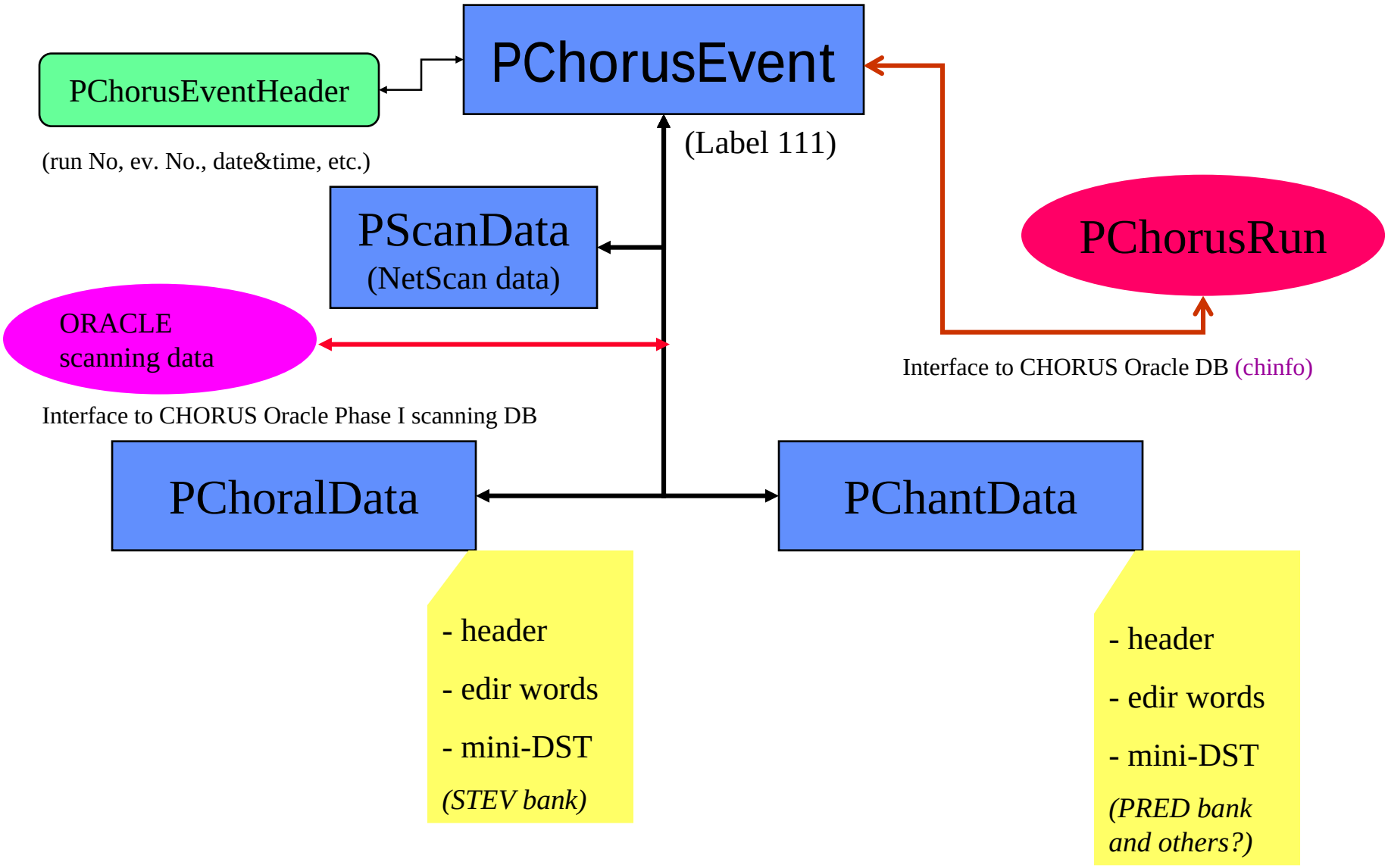
Some (mixing C++ and FORTRAN) code has to be added to the Chorus software infrastructure

- **automatic converter between ZEBRA structure (bank tree) and Objectivity/DB persistent object (C++ class object);**

  (run time mapping is desirable, the kernel of the problem)
  A first (static) try has been made by Jeanne Wilson.
- **Chorus Objectivity/DB schema;**
- **CHANT module for writing its mini-DST in the data base**
- **CHANT module for accessing the data base, reading event from there (both electronic and scanning data - NetScan and others) and constructing ZEBRA structure in the memory.**

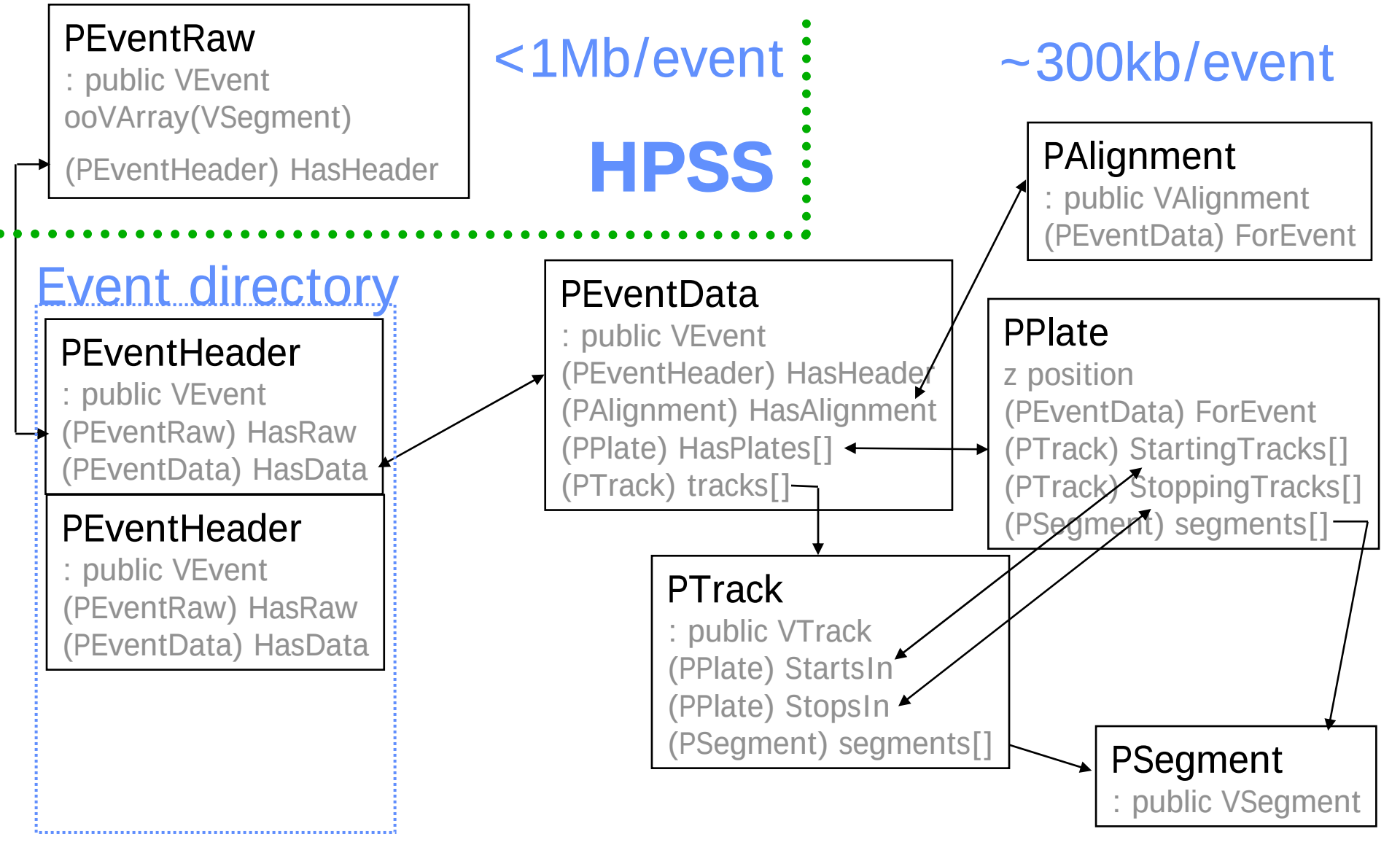
AIM: The end-user will see only a CHANT module, controlled by steering cards.

CHORUS Objectivity/DB persistent objects

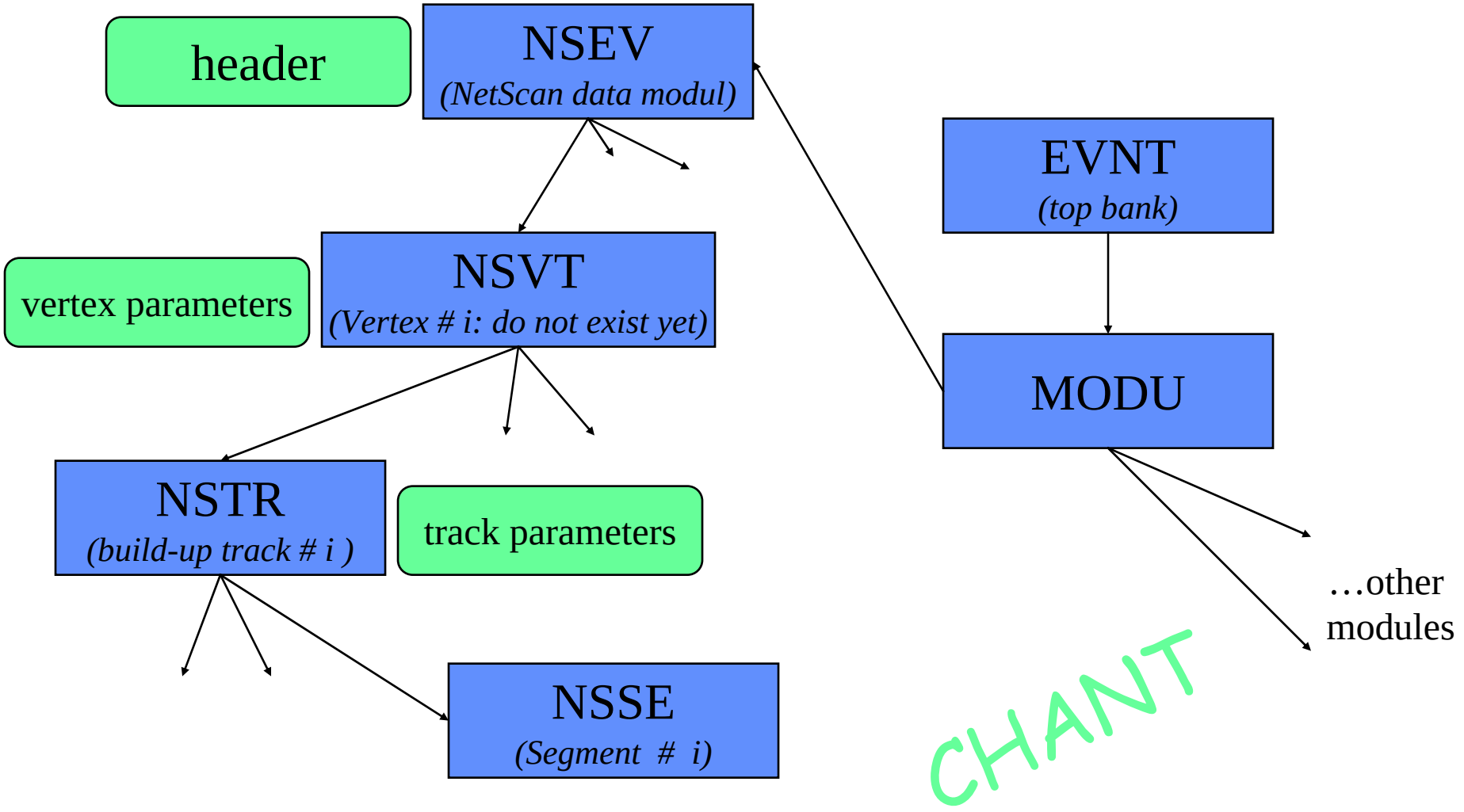


Slide from Bart's presentation on April's Software Workshop

Data model



Proposed ZEBRA bank structure of NetScan data (SCAN module in CHANT)



Stages of realization

- Test model for ZEBRA bank tree ↔ C++ object conversion and storing in Objectivity/DB
- Setting up test DB : CERN NetScan data base + Choral banks for the same events + SCAN module of CHANT
- Setting up Chorus DB with all NetScan and all electronic data
- Setting Objectivity/DB association to data, stored in ORACLE scanning DB and access them inside CHANT SCAN module (an attempt)
- Final DB + access tools