From the Demographics Committee

Usha Mallik, Mike Ronan*, Helen Quinn, Maury Tigner, Jonathan Bagger

* Deceased

A Short History (1)

- Need for career track trend and statistics
- Each important sub-committee formed needed to re-create the info
- Committee formed by Mallik in 2002 (with Frank Sciulli, Mike Barnett and Rocky Kolb in addition (previous members)
- Sciulli's suggestion => to use the PDG database

Continued (2)

- HEPAP and DPF took ownership of the committee
- In 2003 funding to PDG for this cancelled
- Later after appeal from HEPAP chair (Fred Gilman) funding restored
- 2004 - 2005 => responsibilities for the database transferred from Mike Barnett to Mike Ronan

Continued (3)

- Mike Ronan established modern and flexible database software

- Good communication between committee and Mike Ronan with interim cross-checks (Mallik)

- New suggestions incorporated into the database => improved overall quality

Steps toward "Tracking"

- Committee suggests adding a unique ID to each individual for tracking
- Archival data base should not change (PDG)
- Create a modified database for analysis purpose correcting obvious mistakes and adding extra info for tracking-ID
- Tests on small piece OK
- Committee had planned to send an appeal to community along with the NSF-DOE questionnaire
- But……
- Meeting planned for October 06 cancelled

Database and Analysis

- Each year the new data collected is a separate database (was sent by Mike Ronan)

- All previous data into Archival database (also forwarded by Mike)

- Compare both for internal consistency: make multiple passes to correct mistakes and create files with unique identifier (for each individual)

Analysis Strategy (1)

- Main pieces of information used

  Start with :

  – 1) Their surname

  – 2) Their first name (and initials or middle name, if available)

  – 3) The year of their PhD

  These are the same, carried over from institution to institution and (often) unique

Analysis Strategy (2)

- The basic approach is :

1) Read in the main database line by line

2) Fix obvious problems (see examples)

3) Assign a unique number to each individual

4) Look for the entries belonging to each individual

5) Optionally, check for further potential problems

6) Write out an updated database where each individual is "tagged" with their unique number.

Error-checker's printout

1) They have the same surname and the same non-zero PhD year, but different first names.

   e.g. "Smith, James (2000)" vs "Smith, James Q. (2000)"

   e.g. "Smith, James (2000)" vs "Smith, Jim (2000)"

2) They have the same surname and first name, but different PhD years.

   e.g. "Smith, James (2000)" vs "Smith, James (1999)"

   e.g. "Smith, James (2000)" vs "Smith, James (0)"

3) They have the same first name and non-zero PhD year and their surnames are similar (first three characters identical)

   e.g. "Smith, James (2000)" vs "Smithe, James (2000)"

4) They have the same surname and PhD year and their first names are similar (first three characters identical)

   e.g. "Smith, James (0)" vs "Smith, James Q. (2000)"

5) Their first names are similar and their last names are similar, and they have the same PhD year.

   e.g. "Smith, James Quincy (0)" vs "Smithe, James Q. (0)"

Examples from control files

VanGemmeren|Peter|1998|Van Gemmeren|Peter|1998

Kribbs|Graham|1998|Kribs|Graham|1998

Burkiham|Piyabut|2005|Burikham|Piyabut|2005

Cronin-Hennessey|Daniel|1997|Cronin-Hennessy|Daniel|1997

Holstrom|Timothy|2003|Holmstrom|Timothy|2003

Djurcic|Zelimir|2005|Djurcic|Zelimir|2004

Servant|Geraldine|0|Servant|Geraldine|2001

Morrissey|David|0|Morrissey|David|2005

Lillie|Ben|2005|Lillie|Benjamin H.|2005

End-product

- Finally, they are looked over and compared manually.
- Now real analysis for tracking individuals can begin.

Immediate Goal:

Get a first set of analysis results in one year (was hoped for around now)

A cross-check with the new database keeper necessary

Current Status

- Bill Carithers at LBL with professional help will continue….currently catching up
- A lot of progress has been made (as a group)
- Maintaining continuity critical: to pick up from where it was left
- Currently, the project to produce the "tagged" database is well underway
- The committee will soon meet with Bill Carithers