# Data Quality

## Presentation to eRA Project Team

## March 23, 2004

James D. Tucker
Chief, Systems Quality Assurance
and Interfaces Branch

# Definition of Data Quality

Adapted from Dr. Joseph Juran:

Data are of high quality if they are fit for their intended uses by customers in operations, decision making, and in planning.

# We can agree on the following:

- Prevent Errors at the Source

- Management Accountability

# Person Profiles in eRA

## Statistics

| | |
|---|---|
| Profiles | 388,652 |
| Project Persons | 1,538,526 |
| CMPR | 306,652 |
| FEDP | 7,688 |

# Commons Statistics

## Since 01/01/2004

- Total Accounts Initiated      10,171

- Account Verified
  and Accepted      7,732

- Accounts Accepted      7,676

# Commons Registrations

- Accounts Initiated week of 3/14/04    830

- Institution Registrations  03/2004      132

- Institution Registrations  03/22/2004   21

# Identification of Correct Data

Of the 388,652 Profiles

- 69% have valid looking SSNs

- 63% have valid looking DOB

- 33% have missing Middle Initial or Name

- 1% have missing First Names or Initials

# Other Data Issues

- Over 500 Helpdesk since 01/01/2004

  - Profile/Role issues other than Commons

  - CANs, EINs

  - Type 5, 6, 7 creation issues

  - Subprojects

# Summary

- Database is growing

- Identification of Unique Individuals

- System wide approach

- Ensure data quality during correction process

- Thoroughness of Requirements