1994

Age 60 project : consolidated database implementation

Regina Marie Harris
Lehigh University

AUTHOR:

Harris, Regina Marie

TITLE:

Age 60 Project
Consolidated Database
Implementation

DATE: May 29, 1994

Age 60 Project
Consolidated Database Implementation

by

Regina Marie Harris

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Computer Science

Lehigh University

April 15, 1994

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science.

March 23, 1994

Date

Thesis Advisor

Chairperson of Department

ACKNOWLEDGEMENTS

This research was supported by the FAA Civil Aeromedical Institute (CAMI) in Oklahoma City, Oklahoma, under contract number DTFA-02-90-90125.

The author acknowledges a number of individuals for their assistance during this research. First and foremost, I thank my advisor Dr. Donald Hillman for providing overall guidance during the project and being an invaluable resource about database design and implementation.

I thank the following individuals from Lehigh University: Dr. Edwin Kay for providing an understanding of the statistical analysis requirements; Dr. Diane Hyland for contributing overall project guidance; Robert Voros for developing the C and Oracle software; Tom Smull for his expertise in COBOL programming; and Steve Corbesero for this assistance with computer system administration.

I also acknowledge the contribution of James Deimler and Albert Zalcmann from Hilton Systems, Inc. who allowed me to pursue graduate studies despite the disruptions to my work schedule.

I also recognize the Project's Panel of Scientific Experts for assistance and advice during the entire course of this work: Susan Baker, Ph. D., Johns Hopkins University; Donald Campbell, Ph. D., Lehigh University; James Danaher, National Transportation Safety Board; Russell Raymond, M.D., Aerospace Medical Association; Ilene Siegler, Ph.D., Duke University; and Martha Williams, Ph.D., University of Illinois.

In addition, I thank the CAMI contract monitor, Pam Della Rocco, for invaluable guidance, assistance, and understanding throughout the duration of the Age 60 project.

Table of Contents

Table of Contents

List of Figures

List of Tables

Abstract

This thesis describes the development of a Consolidated Database (CDB) for the Federal Aviation Administration (FAA) Civil Aeromedical Institute that incorporated several NTSB National Transportation Safety Board (NTSB) and FAA databases for the evaluation of the Age 60 rule and analysis of aviation accidents. The CDB was designed to link together the various historical databases into a single relationally structured database capable of supporting rigorous data analyses for exploring relationships between age and performance and identifying potential age-sensitive contributing factors. Each source database contained data on a particular aspect of available information including the results of periodic medical examinations required for all pilots, pilot certifications, accident investigations, pilot incidents, and deviations. These source databases utilized diverse database and file management systems including Cobol Indexed Sequential files, ADABAS, System 2000, and Dbase IV. The databases range in size and complexity from 15 million medical examination records in the FAA medical database, each containing approximately 60 elements, to 90,000 accident investigations records, each containing over 6,000 elements.

Since the primary goal of the CDB was to develop a global structure that would integrate the information contained in the source databases, the conceptual and logical design phases of the CDB effort involved identifying semantically similar information that was stored using a variety of formats and data models, specifying consistent formatting and coding schemes, and developing a structure that unified the source database information. The CDB was developed using the Oracle relational Database Management System running on an HP 9000/750 Unix workstation. The resulting CDB contains approximately 10 million records organized into 50 tables with 1,300 columns of information and requires 14 Gigabytes of data storage.

1.0 INTRODUCTION

1.1 Background

Part 121 of the Federal Aviation Regulations (121.383c) contains a provision for mandatory retirement at age 60 of commercial airline pilots-in-command and co-pilots. The rule was established in 1959 with the objective of ensuring continued aviation safety. The rationale behind the rule was that the increased speeds and passenger loads of commercial airliners placed greater demands on pilots with respect to their physical fitness and piloting skills. In addition, existing studies reinforced a belief that progressive deterioration of physiological and psychological functions regularly accompanied increasing age. Other studies raised the concern that sudden incapacitation could be brought on by such events as heart attacks or strokes, which occur with greater frequency among older members of the general population. Although not all individuals experience equivalent age-related deterioration in health and performance, it was nevertheless concluded that an age 60 limitation was prudent on the grounds that performance decrements could not be reliably and objectively measured or predicted on an individual pilot basis.

In 1979, the U.S. Congress enacted P.L. 96-171, an Act to require a study of the desirability of mandatory age retirement for certain pilots and for other purposes. The legislation required the Director of the National Institutes of Health, in consultation with the Secretary of Transportation, to conduct a study to determine the effect of aging on the ability of individuals to perform the duties of pilots with the highest level of safety. The National Institute on Aging was assigned the primary responsibility for implementing the legislation and formed an inter-institute committee to prepare the federally mandated report. The committee decided in favor of awarding a contract to the Institute of Medicine (IOM) of the National Academy of Sciences to provide an objective examination, summary, and assessment of existing scientific knowledge relevant to the questions posed in P.L. 96-171.

The National Institute on Aging established a Panel on the Experienced Pilots Study in 1981 to assist in reviewing the IOM report (Institute of Medicine, 1981). The panel held three public meetings to receive comments on the IOM report and to hear oral statements from several organizations. The panel was unable to identify a medical or performance assessment system that could determine which pilots would pose a safety hazard because of early or impending deterioration in health or performance. It recommended that the Age 60 rule be retained, while at the same time recommending that a systematic collection of medical and performance data be carried out for future consideration of the Age 60 rule. This recommendation forms the basis for the present project.

1.2 Objectives

A principal goal of the Age 60 project was to develop a Consolidated Database (CDB) based upon information contained in several pertinent FAA and NTSB databases in order to facilitate analyses that explore relationships between age and performance and identify potential age-sensitive contributing factors. The ultimate long-term aim of the Age 60 Project is to enhance aviation safety by increasing understanding about the relationships among pilot age, experience, and accident rates.

This thesis summarizes the design, development, and implementation of a CDB to support diverse research needs in the aviation safety and human factors research area and provide more direct access to the wealth of information collected by the FAA and NTSB. It is based upon the project report that described the CDB implementation (Harris, Hillman, and Voros, 1992a). The CDB merges several relevant NTSB and FAA historical databases that are currently organized in heterogeneous formats into a single relationally structured database capable of supporting rigorous data analysis. Each of the historic databases contain information on a particular aspect of available information including the results of periodic medical examinations required for all pilots, pilot certifications and ratings information, accident investigations, pilot incidents, and pilot deviations. A key design goal of the CDB was to allow the information contained in the various historical databases to be linked together to enhance the ability to directly associate these information sources. This consolidation and linkage of these databases allows researchers to pose questions that cannot be readily determined from a single database. Each of these historic databases was developed and is maintained by separate organizational entities. The contents and structure of these databases are based upon the needs of the specific organization. The individual databases are based upon diverse file and database management systems utilizing a variety of hardware and software environments. Semantically similar information is often stored using a variety of formats and data models and each database uses different naming conventions. Researcher access to these historic databases to explore new hypotheses is difficult, if not impossible. The resulting CDB was extensively explored to analyze relationships between age and performance and to identify potential age-sensitive contributing factors. These database experiments included the following:

- Regeneration and expansion of data analyses performed in identified relevant studies,
- Provide answers to questions that are raised in the development of a performance methodology such as suggesting measures and their validation,
- Confirm validation of the identified measures,
- Perform searches that analyze hypotheses on the relationship between age and performance, and
- Perform statistical analyses.

For additional details about the database experiments, the reader is referred to the document — Consolidated Database Experiments (Kay, Harris, Voros, Hillman, Hyland, and Deimler, 1992).

Two companion documents are also available for the CDB:

1. The CDB User's Guide (Harris, Hillman, and Voros, 1992b) provides information on utilizing the CDB to retrieve information.
2. The CDB Programmer's Reference (Harris, Voros, and Hillman, 1992) contains information on the programs used to load information into the CDB and process queries generated by the Intelligent Study Builder.

1.3 CDB Goals and Objectives

The overall goal of the CDB development effort was to provide a flexible system for facilitating the storage, retrieval, and analysis of the full spectrum of information contained in the FAA and NTSB historical databases. The CDB goals were established through discussions with CAMI personnel and elicited from potential users through the use of a questionnaire. More specific goals organized by functional groups are shown in Table 1.

Table 1. CDB Goals

Functional Area	Goals
User Access	Facilitate the ability to obtain access to information that spans the contents of the source databases. Provide the ability for end-users to generate desired ad-hoc queries (i.e., dynamically created one-time use) and reports.
Data Modeling	Provide a common data format that allows the integration of the component databases. Support the complex structure required to manage and manipulate the massive amount of information contained in the source databases.

Table 1. CDB Goals (Continued)

Functional Area	Goals
Query Capability	Allow user to express and retrieve desired information through ad-hoc queries. Increase the user's ability to quickly identify information relevant to their needs and transform this information into the appropriate format. Provide timely information access.
Reporting	Support generation of both standard and custom reports. Allow user-defined organization and aggregation of data. Provide filtering mechanisms for accessing information and selecting relevant information for retrieval.
Future Growth	Provide flexibility to be adaptable for new and changing requirements.
Security	Provide user authorization and password validation. Incorporate security mechanisms that provide the level of security required by Privacy Act considerations.
External Usage	Provide tools for allowing formulation of queries by end-users on a microcomputer to be transmitted to CAMI for review and processing.
Integration with Software Packages	Provide mechanism for integration with standard business programs such as word processors and spreadsheets and statistical software packages.

The information contained in the source databases falls under the provisions of the Secrecy Act (5USC-552 ref: a[m]) and, therefore, appropriate security measures to limit access to restricted data elements was a key consideration.

The overall approach taken in this project was to iteratively develop the CDB as required to support scientific analysis experiments on the CDB. In the interest of maximizing the efficiency of output during the development effort, a parallel approach was adopted. Since the CAIS Medical Database and the NTSB Accident database were the major databases of interest in the database experiments, they were loaded and verified first (as described in Section 6) so that analysis depending on those data could be started while the remaining databases were loaded. The other database of importance in the analysis is the CAIS Certification Database, because it provides a linkage between the accident and medical databases through the pilot's certificate number. Consequently the Certification Database was the third database to be installed in the CDB. The FAA Accident/Incident Database and the Pilot Deviation Database were the final databases to be installed.

1.4 Technical Approach

The CDB was designed to provide a central repository of aviation information and a better tool for use in accessing and maintaining a database that provides sophisticated reporting and analytical capabilities. The CDB permits the linkage of various aspects of pilot information related to associated events such as medical examinations, certifications, accidents, deviations, and incidents. The CDB not only facilitates the storage, retrieval, and analysis of this information, but also increases the user's ability to quickly identify information relevant to their application needs. The capabilities provided by the CDB were designed to facilitate access to desired information and allow users to gain insight into and to keep track of critical factors in a timely and accurate fashion. Key capabilities to be provided by the CDB include:

- Permit the design and creation of a database system for describing the structure and data required to represent the FAA and NTSB's aviation information.
- Facilitate the ability of the user population to effectively utilize the CDB to obtain information needed to fulfill the types of queries required to support their research interests.
- Provide a tool that assists the user in identifying and retrieving needed information and transforming the selected information into the required formats.
- Facilitate the ability of the user to combine and analyze information in a manner that is meaningful to their particular research interest.
- Support the combination of data from any number of information sources in such a way as to provide reporting and analyses capabilities for a wide range of

users and research interests. The CDB permits the combination of data from several historic source databases to produce new insights and more comprehensive results.

- Incorporate mechanisms for data extraction and transformation between the CDB and other software packages such as statistical analysis and spreadsheets.

1.5 Document Overview

The remainder of this document is organized as follows.

- Section 2 describes the source databases included in the CDB.
- Section 3 describes the results of the requirements analysis for the CDB.
- Section 4 describes the computer environment used for the CDB.
- Section 5 presents development of the CDB and discusses the development of the CDB structure.
- Section 6 summarizes the contents of the CDB.
- Section 7 summarizes the method for accessing the CDB.

2.0 FAA AND NTSB HISTORIC DATABASE

The CDB incorporates information from several existing sources, including the FAA Consolidated Airman Information System (CAIS) containing both medical history and certification data, the Accident/Incident Database (AIDS), the Pilot Deviation System (PDS), and the NTSB Accident database. A high level description of the various FAA and NTSB source databases included in the CDB is provided in this section.

2.1 FAA Consolidated Airman Information System (CAIS)

The CAMI Consolidated Airman Information System (CAIS) (CAMI, 1981; CAMI, 1984), maintained in Oklahoma City, contains medical data for pilots as well as airman certificate information. The medical certification system contains information on airman applications for medical certification based upon physical standards prescribed in FAR, Parts 61, 65, 67, and 187. CAIS airman information is organized into the following two groups of information:

1. Pilot certification which contains a single record for each pilot with the most recent information on the pilot's certificates, ratings, accidents, violations, cancellation/revocations, and limitations.
2. Medical certification which contains a record for each certification physical including date of birth, certification class, flight hours (civilian total and last six months), and medical information (cardiovascular, vision, hearing, restrictions, and pathology codes).

Pertinent statistics for CAIS are shown in Table 2. No historic information is available from the pilot certificate database nor is there an indication of when the information for the pilot changes. For example, no information is provided about when a pilot obtained a particular certificate or rating.

Table 2. FAA Consolidated Airman Information System (CAIS) Statistics

Item	Value
Number of records	Approximately 15 million including active and historic records
Approximate number of characters per record	350 medical; 250 airman basic file; 400 airman data (variable due to repeating groups)
Estimated database size for active records	500 Megabytes
Estimated database size for historical records	9.6 Gigabytes
Estimated annual update in records	500,000
Database Administrator	Shirley Dark - medical; Joe Robinson - airman

The CAIS medical system is implemented using the ADABAS data management system. ADABAS is an inverted list system similar to a relational database system but with the following significant differences:

- The data manipulation operators are dependent on the notion of record addressing.
- Referential integrity is the responsibility of the user.

However, ADABAS contains search keys (i.e., combinations of attributes over which indexes can be built). This feature allowed the information contained in the CAIS database to be restructured as a relational system. The CAIS certification information is maintained as COBOL indexed sequential files.

2.2 NTSB Accident Records Database

The NTSB Accident Records Database, maintained in Washington, DC, contains information derived from accidents investigated by the NTSB. Three different forms have been used to collect the data. From 1962 to 1981, an Aircraft Accident Analysis Sheet (NTSB Form 6120.12) based on punched card images was used (NTSB, 1981). In 1982, the form was revised (6120.4D NTSB Accident/Incident Report) to facilitate the collection of data in 693 fields plus a free-form narrative. The data collected from 1983 to the present are based upon the information contained on revised form 6120.4, NTSB Factual Report Aviation, which has 236 fields. The NTSB accident data for 1983 to the present contains an additional 22 files, each of which provides information on a specific area such as a text narrative of the accident, causes factors, copilot information, etc. Information contained in the NTSB Accident Database includes the following: pilot information (age, flight hours for specific aircraft and all aircraft for last 24 hours, 30 days, 90 days, and overall total), weather, cabin crew and passenger information, aircraft information, operational phase, investigation results, and causes and factors. Currently, the NTSB information incorporated into the CDB includes the core file containing material provided on form 6120.4, the accident narrative file, cause factors detailed in the sequence of events file, and the co-pilot file. The database statistics shown in Table 3 are based upon the information supplied by the NTSB on database contents dating back to 1983.

Table 3. NTSB Accident Records Database Statistics

Item	Value
Estimated number of relevant records	375,000 (includes core and supplemental files)
Approximate number of total characters per record	13,737
Estimated relevant number of characters per record	5,000
Estimated database size (based upon relevant information)	1.6 Gigabytes
Estimated annual update	3,000 accidents
Database Administrator	Stanley Smith

2.3 Accident/Incident Database System (AIDS)

The Accident/Incident Database is maintained in Oklahoma City by the FAA's Aviation Standards Office and contains data for general aviation accidents and incidents, air carrier incidents, and air carrier accidents (Boeing Computer Services, 1986). The data in AIDS is available for 1982 to the present. Pertinent information includes:

accident/incident location and time, pilot data including certification code and age, flight hours (total and last 90 days for all aircraft types and specific to aircraft make and model), aircraft data, investigation and remedial actions, fatalities and injuries, causes and factors, and weather conditions. Statistics for the AIDS database are shown in Table 4.

Table 4. Accident/Incident Database System (AIDS) Statistics

Item	Value
Number of records	32,000
Approximate number of characters per record	826
Estimated database size	30 Megabytes
Estimated annual update in records	Not available
Database Administrator	Jack Price

The AIDS Database is implemented using System 2000, which is a hierarchical system. The crucial difference between a hierarchical structure and an equivalent relational structure is that, in hierarchies, information that would be represented in relational databases by foreign keys is represented instead as child-parent links in tree structures. This means that there are inherent difficulties in attempting to provide a relational equivalent to System 2000. However, a relational equivalent can be constructed as a tree to represent the System 2000 structure.

2.4 Pilot Deviation System (PDS) Database

The Pilot Deviation System (PDS) database, maintained in Washington, DC by the Office of Safety Analysis of the FAA (FU Associates, 1989), contains detailed information on reported abnormal flight incidents. Information of interest includes: accident/incident description, aircraft involved, operation phase, weather conditions, flight plan, investigation results, pilot information (date of birth, make/model, specific flight hours — total and last 90 days). PDS is implemented using Dbase IV. Statistics for PDS are shown in Table 5.

Table 5. Pilot Deviation System (PDS) Statistics

Item	Value
Number of records	65,000
Approximate number of characters per record	350
Estimated database size	23 Megabytes
Estimated annual update in records	Not available
Database Administrator	Sarah Hodges-Austin

3.0 CDB REQUIREMENTS ANALYSIS

The first step in the CDB database effort was to perform a requirements analysis starting with the formulation of the goals and objectives of the CDB as described in paragraph 1.3. The user population was surveyed to determine their anticipated usage of the CDB and determine their skills and characteristics. The user surveying approach and results are presented in paragraph 3.1. Next, the functional characteristics of the database and the informational needs of the targeted user population were determined as described in paragraphs 3.2 and 3.3. This approach focuses on the front-end phases of requirements analysis and design because these phases are the basis for the subsequent implementation phases and are, therefore, critical to ensuring the effectiveness of the database system.

3.1 User Survey

An important part of this effort was to determine the various user populations and elicit their specific database needs and requirements for not only the data itself, but for the use of the CDB. For example, the needs of the users who are tasked with investigating specific accidents were much more detailed than researchers who wanted to obtain summary data or view trends. The investigator may want to obtain the detailed medical history of specific individuals as well as their involvement in previous accidents, incidents, and deviations whereas the researcher may want only to determine the overall percentage of pilots meeting specific criteria. These two user classes had distinct needs for accessing, manipulating, and reporting the database contents. Another factor was that the CDB contains a wealth of information and the usage of the database contents to support other research efforts, such as retirement patterns, pilot attrition, evaluation of proficiency tests, etc. required analysis. These users required extensive ad-hoc querying capabilities and, perhaps, integration with statistical analysis packages.

In order to determine the characteristics and requirements of potential CDB users, a CDB questionnaire was developed and distributed both to CAMI personnel and the members of the project's Scientific Panel of Experts. This group represented a spectrum of potential users and diverse research interests. The CDB questionnaire and tabulation is contained in Appendix A. Six responses were obtained and analyzed. Table 6 summarizes the key results of the questionnaire responses.

Table 6. Summary of CDB Questionnaire Result

Functional Area	Results
User Characteristics	Little previous exposure to database management systems. Minimal experience with statistical packages (except SPSS). Prefer microcomputers. Tend to rely on other personnel (primarily programmers) to formulate and process the queries required to extract the desired information.
CDB Usage	Data in CDB collected for specific purposes that are rarely the same as the objectives of the research analyses of the same data. Types of applications to be supported are different from standard types of database queries.
CDB Requirements	Ease of query browsing. Speed and report formatting. Access. Variability of required response times from instantaneous to less than 15 minutes.
CDB User Aids	A glossary of database elements would greatly assist in identifying relevant CDB elements. Prefer hands-on training. Would like a hard copy manual.
Potential CDB Uses	Longitudinal studies of health status of pilots. Types of pathologies. Searches by type of certificate and rating. Indication of groups of pilots with certain characteristics and history or accident involvement. "Core of the National Medical Accident System." Detailed analysis of aging data.Teaching aid in development psychology. Trend analysis (e.g., Accidents with respect to medication use).

3.2 User Characteristics

Based upon the information obtained from the CDB questionnaires and discussions with CAMI personnel, the potential CDB user population was subdivided into the classifications shown in Table 7 with a brief description of their associated skills and knowledge.

Table 7. CDB User Characteristics

User Category	User Characteristics
End Users	Primary usage is eliciting information and generating reports, e.g., ad-hoc queries, link to statistical packages, studies requiring download of information. Primary concern is ease of use. Limited computer knowledge is required. No knowledge of database management systems, operating systems, and programming languages is required.
Database Administrator	Primary usage is incorporating historical database updates into the CDB on a periodic basis. Primary concern is ease of use and facilitation of database maintenance tasks. Supports data integrity and access controls. Maintains data structure. Provides database maintenance tasks. Optimizes performance. Computer knowledge is required. Knowledge of database management systems, operating systems, networks, and programming languages is required.
Developers	Primary usage is developing CDB and associated programs for loading, maintaining, and accessing CDB. Extensive knowledge of programming languages, operating systems, networks and database management systems is required. Extensive software engineering knowledge is required. Primary concern is adequacy of development, debugging, and documentation tools.

3.3 CDB Functional Requirements

This section describes the CDB system requirements to facilitate storage and retrieval of pilot, medical, and performance data in a structured, organized form that supports rigorous data analysis. These CDB requirements are based upon the information collected from a series of interviews and discussions with CAMI personnel and responses to the CDB questionnaire.

3.3.1 Database Management Capabilities

Effective data management is critical to CDB development and to readily provide the user population with the ability to access and manipulate the extensive data contents of the CDB. The CDB database management capabilities are summarized in Table 8. The information contained in the source FAA and NTSB databases will evolve over time, so the flexibility provided by the data management system to meet new requirements is crucial. For example, the NTSB accident database is scheduled for a major redesign in the near future which will eliminate approximately 2,000 of the current elements. In addition, other pertinent database systems may be identified in the future that should be integrated into the CDB.

Table 8. CDB Database Management Capabilities

Category	Requirements
General Capabilities	Support the ability to perform the standard database operations including create table structure, insert, update, delete, and query. Allow access through both an interactive, declarative method for end users and an embedded, procedural method in a high-level programming language for system developers and system database administrators. Support query-intensive operations and organization for searching and rapid information retrieval versus update-intensive with high-volume transaction production. Provide reasonable response time for querying. Instantaneous results are not required. Facilitate use by all levels of users including novices and experts. Represent state-of-the-art Database Management System (DBMS) technology including support for recognized industry standards such as Structured Query Language (SQL).

Table 8. CDB Database Management Capabilities (Continued)

Category	Requirements
Architecture	Support linkage of several historical databases. Support comprehensive query, manipulation, and study of massive amounts of data. Support structured data types. Support a variety of data structuring techniques (e.g., indexing). Permit establishment of security controls to prohibit unauthorized data access and updates at several levels for individual users and groups. Provide independence from logical and physical database structures. Provide tools for optimizing database.
Flexibility	Permit evolution of database contents. Support standard access and interface protocols. Permit future multi-user and remote access capabilities.

The CDB needs to provide the full functionality afforded by a traditional DBMS; however, the ability of end users to readily access and utilize the capabilities afforded by the DBMS is critical. That is, the DBMS must provide the access to the functions in a form tailored to facilitate the ability of end users to input data and readily identify the appropriate information for retrieval and reporting. The CDB also needs to provide schema management capabilities that allow system administrators to define and evolve the data structure, or schema. The ability to modify the database schema through use of a Data Definition Language (DDL) is required for development, maintenance, and extension of the database definition. As the CDB evolves, the capability to define new structural components and extend existing definitions based upon the existing structure is required. Standard database administration utilities are also needed, including restart or recovery, access control and security, performance optimization, disk management, and directory management.

The CDB also must provide capabilities for specifying security controls to prohibit unauthorized data access such as password controls for both user and database access.

Mechanisms are also required to limit the ability for users to update the database structure to the database administrator and update the database contents to the appropriate users responsible for the technology program. The ability to define access level controls to the element level is also desirable to prevent unauthorized access and update capabilities.

3.3.2 Functional Capabilities

The functional capabilities provided by the CDB are shown in Table 9.

Table 9. CDB Functional Capabilities

Category	Requirements
Input Capabilities	Initial database load using DBMS utilities and custom developed routines to transform historic database contents to CDB format including verification and procedures for handling anomalies. Periodic updates from historic databases using DBMS utilities and custom developed load programs. No end user input or updates.
Access/Data Manipulation Capabilities	Via query language and/or report writer for query and data retrieval. Via high-level languages, e.g., C and Fortran, for custom routines and programs. Utilize Graphical User Interface (GUI) for database functions. Provide security and privacy mechanisms to prevent unauthorized access.
Output Capabilities	Provide both tabular and graphic capabilities. Support export of data in formats for use by other software packages such as spreadsheets, word processors, statistical packages, etc. Support generation of both on-screen and printed reports. Provide functions for aggregating data such as counts, sums, minimums, maximums, etc.
Software Development Capabilities	Provide a comprehensive and integrated command set and development tools including screen generator, data dictionary, debugger, and query language. Provide ability to integrate with a programming language such as C or Fortran.
Data Integrity/Quality Control	Automated detection and clear reporting of data entry errors such as incorrect data types, out of range values, etc. Provide ability to cross check information across data fields. Provide programmable features for detection and handling of inconsistencies, null values and data gaps.

Table 9. CDB Functional Capabilities (Continued)

Category	Requirements
Querying Capabilities	Support both ad-hoc querying and generation of standard data reports. Allow users to specify a variety of tabular and graphical formats for query results and ASCII files for use by external programs. Allows database contents to be viewed based upon a user's perspective. Perform complex querying to handle volume of data. Ability to relate data in a variety of ways and condense large quantity of data into few important trends. Ability to extract data across several groups of information. Accommodate the ability to relate data in a variety of ways and support various levels of abstraction and detail. Provide the ability to generate summary statistics about database components. Support the ability to specify joins and views from multiple data sources. Support query optimization capabilities to ensure efficient query operations and minimize response time.
Privacy/Security	Support access restrictions based upon Level II Sensitive Data covered under Privacy act. Precludes remote access except under stringent conditions and cryptographic protection in accordance with FAA regulations. Provide password controls for both user and database access. Provide access level controls to element level to prevent unauthorized access and update capabilities.
Connectivity	Single user, stand alone system with no remote access during current project. Design to permit remote access in future under guidelines and regulations required for handling Level II Sensitive Data covered under Privacy act.

4.0 COMPUTER ENVIRONMENT

A critical aspect of the CDB development was defining the hardware and software environment conducive for both system development and refinement and actual usage by the end user population. Due to the considerations of the Privacy Act, the selected workstation was required to operate as a single-user workstation. The configuration of the suitably powerful hardware configuration involved analysis of the storage requirements to handle the anticipated volume of information required for the source databases, memory requirements, hard-disk performance, and processing speed. The performance of the hard-disk system is critical in ensuring optimum response time. This configuration requires careful consideration since the performance of a database application is to a large measure a function of the power and resources provided by the workstation.

CAMI needs and constraints for the selection of the hardware and software environment were discussed in detail at Project Review Meetings. The purpose of the discussion was to merge CAMI needs with the technically driven requirements of the database development effort. CAMI concerns include the following:

1. The CDB must be operated and maintained by existing data processing staff available at CAMI, and
2. The CDB should not be implemented on the existing VAX cluster.

An analysis of the required computer capabilities along with a description of candidate hardware and software components and recommendations is contained in the Age 60 Rule Study Hardware and Software Evaluation (Hilton Systems, Inc., 1991).

4.1 Hardware Architecture

The type of hardware/software architecture required for the CDB and associated tasks falls in the workstation category. This type of computing is rapidly becoming the architecture of choice in applications featuring the assimilation and integration of data from diverse and logically different sources and the servicing of many different types of user needs. Reduced Instruction Set Computing (RISC), multitasking, graphical user interfaces (GUI), and networking are provided by workstations. Of the major systems examined (Sun SparcStation2, Dec Workstation 5000 Model 200), the Hewlett-Packard (HP) 9000 Model 750 provides superior performance capabilities in the price range and emerged as the clearly superior choice in terms of the following criteria:

- Speed of processor — 66-MHz PA-RISC chip operating at 76 MIPS,
- Amount of memory from 16 to 256 Mb,
- Disk capacity of up to 40 Gb,
- Visual User Environment — user friendly front-end to Unix operating system,
- SCSI-II disk support as well as 4 EISA expansion slots, and
- Support for industry standards such as OSF/Motif GUI.

The selected configuration for the HP 9000 Model 750 is shown in Table 10.

Table 10. HP 9000 Model 750 Configuration

Feature	HP 9000 Model 750
Operating System	HP-UX 8.05 Unix
Processor	PA-RISC
Performance	Clock Speed: 66 MHz SPECmark Rating: 72.2 MIPS: 76
Memory	64 Mb (expandable to 192 Mb)
Color Monitor	19" with 1280 x 1024 resolution
Keyboard	109 keys
Mouse	3-button
Disk Interface	SCSI-II with fast/differential
Disk Capacity	2.6 Gb (internal) with 3 external 3.9 Gb for a total of 14.3 Gb
Maximum Disk Capacity	40 Gb
Floppy Drive	3.5" 1.4 Mb
Backup Storage	1.3 Gb DAT
Networks Supported	NCS, NFS, TCP/IP, BSD Network Services, ARPA Services

Table 10. HP 9000 Model 750 Configuration (Continued)

Feature	HP 9000 Model 750
Graphical User Interface	HP-VUE, OSF/Motif, X-11 Windows
Languages	C, C++, FORTRAN, Pascal

4.2 Software Components

The software components required for the development of the CDB include:

- HP-UX — HP's UNIX operating system.
- HP-VUE — HP front-end tool to facilitate use of the UNIX operating system.
- SPSS — statistical analysis package.
- HP Interface Architect — OSF/Motif GUI development toolkit.
- C — high-level program language for development.

In addition to the above components, selecting the database management system required careful consideration and is described below.

4.2.1 Database Management System

The selection of the data model used to represent information is crucial in order to optimize the capability of the CDB to be organized to provide flexible access to the wealth of information contained in the historical databases. A data model provides the conceptual basis for a DBMS. It indicates how the data is organized into a structure and specifies the operations for the creation and manipulation of data contents including retrievals, insertions, deletions, and modification. Since the typical life cycle for software is approximately five years, the choice should be made to ensure that not only are current data modeling, representation, and manipulation capabilities met, but those envisioned for the near future as well. The selection of the DBMS package has long-term implications for the viability of the CDB and, therefore, the selection process considered not only its suitability for meeting today's requirements but also for accommodating anticipated growth and dynamically changing requirements. A DBMS that does not have the potential to grow with an organization's data management and analyses needs may limit the full informational potential of the collected data.

Only commercially available DBMS were considered for the development of the CDB. In addition to the software constraints listed in paragraph 3.3, the following considerations were taken into account in selecting the DBMS software:

- User-computer interface support including GUI.
- Performance capabilities, particularly for database applications involving Unix workstations.
- Conformance with industry standards such as SQL, GUI, etc.
- Ease of devèlopment and software development, debugging, and optimization utilities.
- Security and privacy capabilities.
- Hardware requirements such as storage.
- Acquisition and operational costs.

Based upon our analysis of the desired data management and analytical capability for the CDB, a Relational DBMS (RDBMS) offered the most promise to provide a flexible tool for managing the full spectrum of information required to be incorporated into the CDB. A RDBMS provides the ability to manipulate large quantities of information efficiently, provides a high-level query and reporting mechanism, and incorporates integrity constraints. The SQL languages provided by RDBMS have been specialized for database operations and can accomplish in a few lines of code what would typically require numerous lines in another DBMS. SQL has been designed for efficiently accessing information from both single and multiple tables. It directly supports retrievals that span several tables, e.g., a join of tables on the basis of common information. In addition, SQL incorporates many features that optimize use in a client-server environment such as the ability to store precompiled SQL statements on the server for frequently executed queries.

A brief description of the RDBMS is presented below. Version 7.0 of Oracle, which is in the process of being released, overcomes many of the limitations of Version 6.0, particularly in the areas of optimization. Features of interest in Version 7.0 include cost-based optimization, declarative referential integrity capabilities, and provision for database alerts when specified conditions are encountered.

4.2.1.1 Relational Database Management Systems

E. F. Codd (1970, 1972a, 1972b, 1979) introduced the relational data model in the early 1970's to provide additional flexibility in organizing large databases and correct many of the problems associated with earlier hierarchical and networking data management systems. RDBMS were designed to solve the needs of business data processing applications. A fundamental concept of the relational data model is data independence, i.e., the physical storage of the data is separated from the user's view. The relational data model structures information into tables containing two-dimensional arrays of rows and columns of data elements. Each row represents real-world entities or relationships between a set of values. The columns represent the properties or fields of information that describe the entity. A table is an instance of a relation containing characteristics of fixed types that represent properties of the entities and relationships and a primary key. The primary key is a sequence of columns which uniquely identify each row. The columns of information are specified using predefined data types such as string, integer, real, date, etc.

A big advantage of RDBMS over previously available DBMS is that the user can access the data without any knowledge of the database structure or predefined access paths. Most RDBMS provide querying facilities both from within high-level programs (e.g., C, Fortran) and interactively through a high-level declarative language. SQL has become an ANSI standard database language for defining the table structures and performing retrieval, insertion, deletion, and update operations on the database contents.

The advantages of RDBMS for managing information are summarized below:

- **Data Management** — Provides excellent capabilities for handling standard business data processing data types such as character strings, integers, floating point numbers, date, time, money, etc. and built-in operators (e.g., +, -, *, /).
- **Querying** — Provides the defacto industry standard for query operations, e.g., SQL.
- **Data Model** — Formally based on set theory, relational algebra and first order predicate logic.
- **Performance** — Efficient and well-suited for querying simple data structures and traditional transaction processing. Optimal performance provided for DBMS with relatively few (e.g., less than 200 tables).

4.2.1.2 <u>Oracle</u>

The Oracle RDBMS was selected for the CDB. Oracle provides a comprehensive set of programming tools and is SQL-based. It has an excellent track record for use in the management of large databases. The salient characteristics of the Oracle RDBMS are shown in Table 11.

Table 11. Oracle Version 6.0 Features

Feature	Oracle Version 6.0
Hardware Platforms	IBM, DEC, Sun, HP, Data General, IBM PC, Mac PC
Operating Systems	Unix, Ultrix, VMS, DOS, Apple Finder
Standards	OSF ANSI (SQL)
Data Organization	Relational DBMS
Host Language Interfaces	C, Cobol, Fortran
Languages Calls	From within language or embedded
Query Method	Commands, query-by-forms, SQL ANSI Levels 1 and 2
Application Generator	Yes
Embedded SQL Support	Yes
Menu generator	Yes

Table 11. Oracle Version 6.0 Features (Continued)

Feature	Oracle Version 6.0
Supported Data types	Character, numeric (double precision, decimal, small and long integer), date, time, long, raw, long raw (raw is used for graphics)
Number of characters per field	256 data/64 Kb long data
Maximum number of fields	256 per table
Number of records per file	Limited by disk space
Maximum Number of relations	64,000
Maximum size	4,000 Gb per database
Number of fields allowed per index key	Unlimited
Number of Indexes per file	Unlimited
Programming Tools	Data Dictionary SQL*Forms SQL*Loader SQL*Menu
4GL	SQL*Plus
Spreadsheet	SQL*Calc
Report Writer	SQL*Report Writer
Gateways	SQL*Net SQL*Connect-Vax RMS, DB2, SQL/DS, IMS
Loader utility	SQL*Loader
Maximum query criteria	255
Report generation capabilities	Full-screen report painter, custom reports
Calculated Output fields	Yes
Security levels	Passwords, record/field/file/table level
Integrity	Triggers, stored procedures, and custom developed
Backup/Restore	Journal log, audit trail roll back/roll forward

4.2.2 Software Development

Two types of CDB software are required. The first type of software is that required for CDB maintenance and is used to initially load the CDB and incorporate periodic updates. The primary software utilized for these activities are those provided directly by the DBMS, e.g., Oracle's SQL*Loader. However, due to the nature of the processes required to transform the historical databases into the CDB format and verify their contents, some custom software utilizing a programming language was required.

These custom routines were developed in C using Oracle's Pro*C host language interface since the C programming language provides a powerful and varied set of functions and data structures that can be utilized for the type of data manipulation required for the historic database transformations. The modular capabilities of the C language were utilized to develop well-structured software that can readily be understood and maintained. These custom C routines supply standard operations, e.g., data transformation, and should not require extensive modification or attention in the future.

The second type of software required by the CDB is for developing programs required to support database experiments and access which need more sophisticated capabilities than that provided by SQL*Plus. For example, calculations of accident rates cannot be readily achieved using the SQL capabilities provided in Version 6.0 of Oracle. Frequently, the database experiments require several iterations of the data with slight modifications of the parameters, e.g., pilot age. The amount of time required to obtain information for this type of retrieval can be greatly reduced by accessing all the database records only once and using C array capabilities to store the needed information for each iteration. These programs were also developed using Oracle's Pro*C software.

5.0 CDB DEVELOPMENT

The CDB was developed in an incremental fashion in order to rapidly develop an initial database that would allow data analysis as early in the project as possible. Additionally, the media supplied by the cognizant database administrators for each source database arrived in stages and, in many cases, were delivered initially on unreadable and/or unusable media. The initial database contained a portion of the database contents and was used as a prototype to assess the viability and usability of the design. The prototype was also used to explore additional issues affecting the CDB and to lay the groundwork for subsequent expansion. The contents of the initial database were expanded, first, based upon the needs of the database experiments and, second, based upon the availability of a readable set of media. Each version established precise definitions of the required database entities, their properties, and important relationships between them. These specifications were then utilized to check for consistency and completeness in the specifications and to disambiguate features of the previous prototypes. Each resulting prototype was extensively utilized in the database experiments which assisted in ensuring that the database design reflected both the needs of the current database experiments but also access requirements envisioned based upon discussions with the target user population.

The CDB design and development effort involved the following tasks:

- Developing a detailed understanding of the source database data elements, the relationships among the elements, and the rules for creating, validating, and manipulating the data elements.
- Specifying naming conventions, coding procedures, and linkage mechanisms.
- Grouping the data elements into relations.

- Designing a database structure for the components of the initial CDB prototype.
- Developing an initial CDB prototype for a carefully selected subset of the CAIS medical and NTSB accident databases for the years 1987-1988 required for preliminary data analysis including the creation of the database tables, population of the database, validation of the contents, and development of Pro*C programs to extract the desired information and transform it into the appropriate form for analysis.
- Incrementally expanding the initial CDB prototype to provide data management capabilities for additional source databases and time periods.
- Developing database tables to contain a glossary of database elements.
- Developing Pro*C programs that allows the generic specification of database requests.
- Developing a microcomputer-based tool to allow the specification of database queries to be subsequently executed on the HP workstation.

5.1 RDBMS Terminology

To assist the reader in understanding the concepts presented in this document, a brief description of the standard RDBMS terms is presented. Table 12 illustrates a relational table with the name MEDICAL that contains information representative of that in the FAA CAIS Medical database. Table names are shown in upper case in this document. A table represents a two-dimensional structure consisting of rows and columns of information. Each column contains a basic unit of data used to describe a particular fact or characteristic about the information contained in the table. Each unique set of columns represents a record, or row, of information. Table 12 illustrates a MEDICAL table containing four columns and five rows of information. The columns for the MEDICAL table represent the certification number (Column 1), medical date (Column 2), medical class (Column 3), and age (Column 4). Each column is assigned a unique label referred to as a column name. Column names are shown in lower case with distinct parts of the name separated by an underscore (_), e.g., certification_no, med_date. The fields of data to be stored in each column are specified as a data type such as number, date, or character string. The contents of a column for a particular row are referred to as a value. For example, the value of Column 2 for Row 3 is 02-AUG-89.

Table 12. An Example of a Relational Table

Table Name	Column 1	Column 2	Column 3	Column 4
MEDICAL	certification_no	med_date	med_class	age
Row 1	123456789	01-FEB-89	1	50
Row 2	234567890	01-FEB-89	2	36
Row 3	123456789	02-AUG-89	1	51
Row 4	234567890	01-FEB-90	2	37
Row 5	123456780	03-FEB-90	1	51

Each row in the table is uniquely identified through the assignment of a primary key. The primary key can consist of one or more columns. The certification_no column shown in Table 13 is not unique. However, when paired with the med_date, a unique primary key is formed consisting of both columns. To associate information between tables, a foreign key is required. The certification_no column of the MEDICAL table can be used to link to the ACCIDENT table through use of the pilot_certification_no column as shown in Table 13. This capability would allow a query to be made requesting medical information for a pilot involved in an accident.

Table 13. An Example of a Relational Table Foreign Key

MEDICAL

certification_no	med_date	med_class	age
123456789	01-FEB-89	1	50
234567890	01-FEB-89	2	36
123456789	02-AUG-89	1	51
234567890	01-FEB-90	2	37
123456780	03-FEB-90	1	51

↑
Foreign Key to link with ACCIDENT table ↘

↖ Foreign Key to link with MEDICAL table ↓

ACCIDENT

accident_no	accident_date	flight_conduct	pilot_certification_no
81-12345	01-FEB-81	121	123890123
81-23456	14-FEB-81	135	098703456
81-34567	02-AUG-81	91	234567890
81-45678	01-SEP-81	91	794569700
81-567890	03-OCT-81	91	609482955

5.2 CDB Database Structure Development

As the CDB prototype was expanded, a global structure evolved that unified the diverse database information, identified semantically equivalent data elements, and, where possible and within the limits of available resources, specified consistent formatting and coding schemes. The information contained in each source database was examined and similar information, such as aircraft owner, pilot data, weather, accident causes, etc. was grouped into a set of Oracle tables. The characteristics that describe the data elements were examined along with their acceptable values. In developing the overall structure of the CDB, the following were taken into consideration:

- The coordination of data access by different programs and applications,
- The preservation of data consistency,
- The correctness of the data,

- The methods used to encode data items,
- Provisions for data protection required due to the sensitive nature of the information and Privacy Act requirements,
- The user views of the data,
- The nonredundancy of stored data, and
- Access flexibility.

The development of the overall CDB structure included the following steps:

1. The data entities appropriate for incorporation into the CDB from each source database were defined as well as the relationships between the elements, the characteristics, or attributes, that described the entities, and acceptable values.
2. New or derived data columns were specified where necessary or desirable to facilitate database queries.
3. The schemas required to represent the source database, including detailed specifications of required columns, data types, and size, were specified.
4. Each source database schema was converted to relational form.
5. A global relational schema was devised by synthesizing the local schemas in such a way that successive table structures were compared with previous structures and either merged with already defined base relations or used to establish new relations. This is a standard technique of relational database construction based on normalized structures. The schema was a concise description of the data requirements of the CDB users, including detailed specifications of data types, relationships, and constraints.
6. The local schemas for each individual source database were mapped to the global schema.

Several of the source databases contain a large number of attributes per record. For example, the NTSB 1983 core table contained over 500 data elements and AIDS contains 161 per record. Organizing this information in the CDB required considerable restructuring in order to take advantage of the relational capabilities of the RDBMS. This restructuring involved identification of logically related information to be represented in separate tables. For example, accident data was grouped into categories such as accident, aircraft, airport, pilot, etc. An additional consideration was that Oracle only permits a maximum of 255 columns of information per table.

Information that was repeated a variable number of times, such as abnormalities, EKG defects, pathology codes, and medical restrictions for CAIS medical information, was placed in separate subtables that are associated with the parent table through the use of foreign keys. In this case, the certificate number and social security number were used to link the tables. However, to minimize the processing time required when retrieving information from these subtables, additional columns representing the number of these conditions were added to the appropriate table. For example, the columns num_abnorm, num_restrict, num_ekg_def, num_path_code in the MEDICAL table indicate how many conditions exist. When querying this information, the parent table can be examined to determine if the number of conditions is greater than 0 and in only that case perform the linkage to the subtable to examine the codes.

The source databases frequently contained extensive elements that repeated information for a different entity. For example, the NTSB and PDS source databases maintained separate sets or files of data elements for pilot and co-pilot information. In the CDB, a table was created to represent pilot information and an additional element added to indicate the crew member position as pilot or co-pilot. In this manner, the column names for representing pilot information were identical. This technique to logically group information together minimized storage requirements and facilitated the querying process.

Two separate formats are used for the CAIS medical information -- active and historic. The historic information is a partial subset of the active information. Only relatively recent information is available in the active format; all archival storage is maintained using the historic format. (The transformation of the active data to the historic format results in the loss of certain fields.) The CDB medical information is based upon the elements available in the historic format.

Additionally, provisions for data protection required due to the sensitive nature of the information and to Privacy Act requirements were specified as part of the database design. Oracle provides two mechanisms for achieving the desired levels of security:

1. User identification and password requirements for access to the database.
2. Access restrictions to the tables and views for specific operations, e.g., insert, update, delete, and query.

The CDB user population was categorized into the following primary groups:

1. System Database Administrator (DBA) who will have full read, write, and update access to all database elements,
2. FAA and other users who have the appropriate clearances required by the Privacy Act will have read access to all database elements, and
3. Other users who will have read access to only those database elements that are not covered as part of the Privacy Act.

In consultation with CAMI personnel, the data elements that require limited access were established.

An important aspect of the database design is that it not preclude subsequent studies and analyses. For this reason, the total information content of the individual databases was preserved in the design. This was accomplished by including all relevant attributes in the data definition stage, although only a small fraction of the total collection will ever be used in a given search. The only database elements that were not included were those that are specifically related to the maintenance of the database by the associated database management system. For example, the `Addition Badge

(ADD_BADGE)' element in the PDS system which contains the LOGONID of the operator or the `Quarter' element which is used for counting purposes.

Since one of the key usages of the CDB is to investigate relationships between pilot age and other factors, a column was added to each appropriate database to contain the pilot's age based upon the date of the event -- medical examination, date of accident or deviation, calculated using the SQL date functions TRUNC and MONTHS_BETWEEN as follows:

age = TRUNC(MONTHS_BETWEEN(event_date, date_of_birth)/12, 0)

Many potential queries also matched tables based upon the year of the medical examination or accident. To speed the retrieval process and avoid the requirement for use of specialized date functions, an additional column was added to the tables containing medical and accident data to represent the appropriate year. The addition of these columns of information greatly facilitated the development of queries using age, medical date, and accident date as factors, and minimized the potential for calculation errors.

5.3 Naming Conventions

Consistent naming conventions were established for objects within Oracle such as table names and column names which assists in quickly relating a specific element to a CDB table to an original source database. The naming conventions are based upon Oracle's limit of 30 characters for the name of any database object. English-like names using familiar terms were used wherever possible and the use of abbreviations was minimized. All table names begin with a brief description of the source database name followed by an underscore (_) as follows in Table 14.

Table 14. Table Identifiers

Table Identifier	Source Database
AIDS	Accident/Incident
MED	CAIS Medical
NTSB	NTSB Accident
PDS	Pilot Deviation System
PILOT_CERT	CAIS Pilot Certification

Since the NTSB Accident information is organized in three distinct formats and some information is available only for certain time periods, the following suffixes were used to indicate relevant time periods for NTSB information:

_81 — 1981 and earlier.

_82 — 1982.

_8182 — 1982 and earlier.

The primary table, e.g., which does not include a suffix, includes all data elements defined for the NTSB format used after 1982. If any data elements from the previous formats were identical in semantic meaning, format, and coding schemes to those used after 1982, this information was stored in the 1982 and after format. For example, the ni_cabin_fatal column in the NTSB_INJURY data contains information from the cabin_fatal_inj data elements used after 1982, the cabin_attn_fatal data element used prior to 1982, and the f500_cabin_fatal data elements used in 1982. However, the data element_other_govemnt_fatal was only collected prior to 1982 and was therefore stored in the NTSB_INJURY_81 table. Similarly, f500_instruct_fatal data element was only available for 1982 and was stored in the NTSB_INJURY_82 table. To further complicate the issue, faa_pernsl_fatal data element for the data prior to 1982 and f500_faa_fatal for 1982 data both contain information on the number of FAA personnel with fatal injuries. Since this information is available for both data formats, the column ni_faa_fatal_8182 is contained in the NTSB_INJURY_8182 table.

In addition, unique prefixes were developed for all column names to identify the associated table. These prefixes are contained in the description of the CDB Oracle tables in Section 5.5.

5.4 Coding Schemes

The source databases used a variety of schemes for encoding information. These coding schemes are used to associate an abbreviation or mnemonic with a particular entry in a set of possible values for a column of information. Coding schemes are used to minimize storage requirements, particularly for large character fields, facilitate data entry, and attempt to ensure consistency. An example of a coding scheme is the two character code assigned to each state, e.g., NY for New York. To facilitate the use of the CDB, common encoding schemes were developed for key pieces of data, such as pilot medical certification class, based upon the values assigned in the source databases. Table 15 contains the common coding scheme developed for key CDB information components. All alphanumeric codes are specified in upper case. In addition, information that was coded as Yes, No, or Unknown was converted to the following:

Y -- for Yes entries,

N -- for N entries, and

NULL -- for unknown/invalid information.

The PDS database does not include database elements that required the use of the CDB codes except for the Yes/No/Unknown and is not shown in Table 15.

Table 15. Encoding Schemes CDB, FAA Medical, NTSB, and PDS Databases

	CDB	FAA Medical	NTSB			AIDS
			Prior to 1982	1982	1983 to Present	
Pilot Medical Class						Not used
I	1	11-19	A,D,G	I	2	
II	2	21-29	B,E,H	II	3	
III	3	31-39	C,F,I	III	4	
None	0	0	J	Blank	1	
Unknown	Null	0	Z	Blank	0	
Accident Incident Code		Not used				
Accident	A		A,F	A	1	A, B
Incident	I		C,H	I	2	C, G
Flight Conducted		Not used	Not used			
14 CFR 91	91			91	1	91
14 CFR 91D	91D				2	91
14 CFR 103	103				3	103
14 CFR 105	105				4	105
14 CFR 121	121			121	5	121
14 CFR 123	123			123		123
14 CFR 125	125			125	6	125
14 CFR 127	127			127	7	127
14 CFR 129	129			129	11	129
14 CFR 133	133			133	8	133
14 CFR 135	135			135	9	135
14 CFR 137	137			137	10	137
14 CFR 141	141			141		141
Valid Medical Certificate		Not used				Not used
Expired	EX		G, H, I		4	
No medical certificate	NO		J	4	5	
Not valid medical for this flight	NV			3	3	
Valid medical-no waivers/ limitations	VM		A, B, C	1	1	
Valid medical-with waivers/ limitation	VW		D, E, F	2	2	

Table 15. Encoding Schemes CDB, FAA Medical, NTSB, and PDS Databases (Continued)

	CDB	FAA Medical	NTSB			AIDS
			Prior to 1982	1982	1983 to Present	
Pilot Certificate		Not used				
Airline transport	A		D	ATP	4	1
Airline transport w/FI certificate	AI		G			
Commercial	C		C	COM	3	2
Commercial w/FI Certificate	CI		F	CFI		
Flight engineer	E				6	
Foreign	F		H	FRN	9	
Flight Instructor	I				5	
Military	M			MIL	7	5
None	NO		I	NON	8	1
Private	P		P	PRI	2	3
Private with FI Certificate	PI		E			
Student	S		A	STU	1	4
Special Purpose	SP					
Unknown	UN		Z			6
Pilot Sex			Not used			Not used
Female	F	F		F	2	
Male	M	M		M	1	
Pilot Profession		Not used				Not used
Aircraft mechanic	AM		X		4	
Clergy	B		H		10	
Business	C		C, D, F, N		5	
Doctor/dentist	D		D, K		7	
Engineer	E		J, S		12	
Farmer/rancher	F		E		13	
Government employee	G					
Lawyer	L		B		6	
Miscellaneous civilian	MC		Q, R, T, U, V, W			
Other-military	OM		I		3	

Table 15. Encoding Schemes CDB, FAA Medical, NTSB, and PDS Databases (Continued)

	CDB	FAA Medical	NTSB			AIDS
			Prior to 1982	1982	1983 to Present	
Pilot Profession (Continued)						
Pilot-military	PM				2	
Police	P		L		8	
Pilot-civilian	PC		G		1	
Retired	R				14	
Student	S		P		9	
Salesman	SA					
Teacher	T		M		11	

5.5 Oracle Tables

The Oracle tables associated with each of the source databases are illustrated in Table 16. Table 17 lists the table names, the unique prefixes associated with the column names, a brief description, and an indication of the time period. The 81, 82, and 8182 suffixes are appended to the table names to indicate relevant time periods for NTSB Accident related information.

Table 16. CDB Structure

Source Database	Information	Table Name	Primary Key
CAIS Airman	Pilot	PILOT_Cert_130	P_Certicate_No, C_SSN
		PILOT_Cert_132	P_Certicate_No, P_SSN
		PILOT_Cert_Rate	P_Certicate_No, R_SSN
		PILOT_Cert_ Address	P_Certicate_No, PA_SSN
		PILOT_Cert_Type	P_Certicate_No, R_SSN
CAIS Medical	Medical	Medical	Certificate_No, SSN
		MED_Abnorm	Certificate_No, SSN
		MED_Path_Code	Certificate_No, SSN
		MED_EKG_Def	Certificate_No, SSN
		MED_Restrict	Certificate_No, SSN

Table 16. CDB Structure (Continued)

Source Database	Information	Table Name	Primary Key
NTSB Accident	Accident	NTSB_Accident	NA_NTSB_No, NA_Accident_Date
		NTSB_Accident_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Accident_82	NA_NTSB_No, NA_Accident_Date
		NTSB_Collision_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Ditching_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Aerial_Application_81	NA_NTSB_No, NA_Accident_Date
	Airport	NTSB_Airport	NA_NTSB_No, NA_Accident_Date
		NTSB_Airport_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Airport_82	NA_NTSB_No, NA_Accident_Date
	Injury	NTSB_Injury	NA_NTSB_No, NA_Accident_Date
		NTSB_Injury_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Injury_82	NA_NTSB_No, NA_Accident_Date
		NTSB_Injury_8182	NA_NTSB_No, NA_Accident_Date
	Aircraft	NTSB_Aircraft	NA_NTSB_No, NA_Accident_Date
		NTSB_Aircraft_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Aircraft_82	NA_NTSB_No, NA_Accident_Date
		NTSB_Aircraft_8182	NA_NTSB_No, NA_Accident_Date
		NTSB_Part	NA_NTSB_No, NA_Accident_Date, NP_Part_Name

Table 16. CDB Structure (Continued)

Source Database	Information	Table Name	Primary Key
NTSB_Accideent	Flight	NTSB_Flight	NA_NTSB_No, NA_Accident_Date
		NTSB_Flight_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Flight_82	NA_NTSB_No, NA_Accident_Date
	Weather	NTSB_Weather	NA_NTSB_No, NA_Accident_Date
		NTSB_Weather_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Weather_82	NA_NTSB_No, NA_Accident_Date
	Pilot	NTSB_Pilot	NA_NTSB_No, NA_Accident_Date, NP_Crew_Position
		NTSB_Pilot_81	NA_NTSB_No, NA_Accident_Date, NP_Crew_Position
		NTSB_Pilot_82	NA_NTSB_No, NA_Accident_Date, NP_Crew_Position
	Owner	NTSB_Owner	NA_NTSB_No, NA_Accident_Date
		NTSB_Owner_81	NA_NTSB_No, NA_Accident_Date
		NTSB_Owner_82	NA_NTSB_No, NA_Accident_Date
	Passengers & Crew	NTSB_Passenger	NA_NTSB_No, NA_Accident_Date, NPA_PASS_Name
		NTSB_Crew	NA_NTSB_No, NA_Accident_Date, NC_Crew_Certif_No
	Cause	NTSB_SOE	NA_NTSB_No
		NTSB_SOE_Causes	NA_NTSB_No, NSC_Cause_No
		NTSB_Cause_81	NA_NTSB_No, NA_Accident_Date

Table 16. CDB Structure (Continued)

Source Database	Information	Table Name	Primary Key
NTSB_Accident	Narrative	NTSB_Narrative	NA_NTSB_No, NA_Accident_Date, NN_Narr_No
Accident/Incident	Pilot	AIDS_Pilot	A_AIDS_No, AP_Certificate_No
	Information	AIDS_Info	A_AIDS_No
	Aircraft	AIDS_Aircraft	A_AIDS_No
Pilot Deviation System	Deviation	PDS_Deviation	D_Report_Number
	Pilot	PDS_Pilot	D_Report_Number, DP_Certificate_No

Table 17. Summary of CDB Tables and Prefixes

Table Name	Column Prefix	Description	Number of Columns	Years
AIDS_AIRCRAFT	AA	Aircraft information	36	1986-1991
AIDS_INFO	A	Accident/Incident general information	102	1986-1991
AIDS_PILOT	AP	Pilot information	21	1986-1991
MED_ABNORM	ABNORM	Pilot medical abnormalities	4	All
MED_EKG_DEF	EKG_DEF	Pilot EKG information	4	All
MED_PATH_ CODE	PATH	Pilot medical pathology codes	5	All
MED_RESTRICT	RESTRICT	Pilot medical restrictions	4	All
MEDICAL	M	Pilot medical information	65	All
NSTB_AIRCRAFT	NAC	Aircraft information	45	1983 - present
NSTB_AIRCRAFT _81	NAC	Aircraft information	15	Prior to 1982
NSTB_AIRCRAFT _8182	NAC	Aircraft information	8	Prior to 1983
NSTB_AIRCRAFT _82	NAC	Aircraft information	62	1982
NTSB_ ACCIDENT	NA	General accident information	45	All
NTSB_ ACCIDENT_81	NA	General accident information	84	Prior to 1982

Table 17. Summary of CDB Tables and Prefixes (Continued)

Table Name	Column Prefix	Description	Number of Columns	Years
NTSB_ ACCIDENT_82	NA	General accident information	17	1982
NTSB_AERIAL_ APPLICATION	NAA	Aerial application information	24	Prior to 1982
NTSB_AIRPORT	NAP	Airport information	21	1983 - present
NTSB_AIRPORT_ 81	NAP	Airport information	29	Prior to 1982
NTSB_AIRPORT_ 82	NAP	Airport information	15	1982
NTSB_CAUSE_81	NCA	Cause factors	14	Prior to 1982
NTSB_ COLLISION_81	NC	Collision information	18	Prior to 1982
NTSB_CREW	NC_CREW	Crew member names and addresses	7	All
NTSB_DITCHING	ND	Ditching information	15	Prior to 1982
NTSB_FLIGHT	NF	Parameters associated with the flight	20	All
NTSB_ FLIGHT_81	NF	Parameters associated with the flight	7	Prior to 1982
NTSB_ FLIGHT_82	NF	Parameters associated with the flight	4	1982
NTSB_INJURY	NI	Injury count by category and type of individual	67	All
NTSB_INJURY _81	NI	Injury count by category and type of individual	15	Prior to 1982
NTSB_INJURY _8182	NI	Injury count by category and type of individual	24	Prior to 1983
NTSB_INJURY _82	NI	Injury count by category and type of individual	19	1982
NTSB_ NARRATIVE	NAR	Text description of accident	4	1982- present
NTSB_OWNER	NO	Information on the owner of the aircraft	18	All
NTSB_ OWNER_81	NO	Information on the owner of the aircraft	6	Prior to 1982

Table 17. Summary of CDB Tables and Prefixes (Continued)

Table Name	Column Prefix	Description	Number of Columns	Years
NTSB_ OWNER_82	NO	Information on the owner of the aircraft	5	1982
NTSB_PART	NP	Information associated with part failures	17	1983-present
NTSB_ PASSENGER	NPA	Passenger names and injury codes	4	All
NTSB_PILOT	NPIL	Pilot and copilot information	120	All
NTSB_PILOT_81	NPIL	Pilot and copilot information	17	Prior to 1982
NTSB_PILOT_82	NPIL	Pilot and copilot information	20	1982
NTSB_RECENCY _82	NPR	Recency of various flight operations	9	1982
NTSB_SOE	NS	Information associated with accident causes and factors	4	1982-present
NTSB_SOE _CAUSES	NSC	Details of accident causes and factors	15	1982-present
NTSB_WEATHER	NW	Weather information	38	All
NTSB_WEATHER _81	NW	Weather information	15	Prior to 1982
NTSB_WEATHER _82	NW	Weather information	24	1982
PDS_BASIC_ INFORMATION	D	Basic information associated with deviation	99	1986 - 1991
PDS_PILOT	DP	Information on pilot and co-pilot associated with deviation	17	1986 - 1991
PILOT_CERT_130	C	Basic pilot information	16	All
PILOT_CERT_132	P	Pilot certification information	16	All
PILOT_CERT_ ADDRESS	PA	Pilot address information	12	All
PILOT_CERT _RAT	R	Pilot ratings	3	All
PILOT_CERT _TYPR	R	Pilot types of ratings	3	All
Total Number of Columns			1,324	

A CDB database glossary was developed to describe all the columns associated with each table contained in the CDB and is contained in the CDB User's Guide (Harris et al, 1992).

5.6 Table Linkage

Unique identifiers are required to link the CDB tables based upon the information available from the source databases. For example, a mechanism for associating a pilot involved in an accident with that pilot's medical examination information was needed. The certification, accident (1982 to present), pilot deviation, and accident/incident database information contain the pilot certification number. However, the pilot records in the CAIS Medical Certification database are identified only by the pilot's social security number. Since the CAIS Certificate Database contains both the pilot's certificate number and social security number, this information was used to add an additional data element containing the pilot's certificate number to each medical record for the pilot. This mapping could not be readily accomplished because certificate numbers in the CAIS Certificate Database are not unique and it appears that the certificate numbers are reused. In order to uniquely identify a pilot, the following combination of database elements are required:

- NTSB Accident, Pilot Deviation -- certificate number and date of birth
- FAA Certificate -- certificate number and social security number

It is important to note that the personal identifiers in the NTSB Accident database prior to 1982 have been removed and, therefore, accident information in this period cannot be linked to other databases. Figure 1 illustrates how the CDB tables can be linked through the pilot's certificate number.

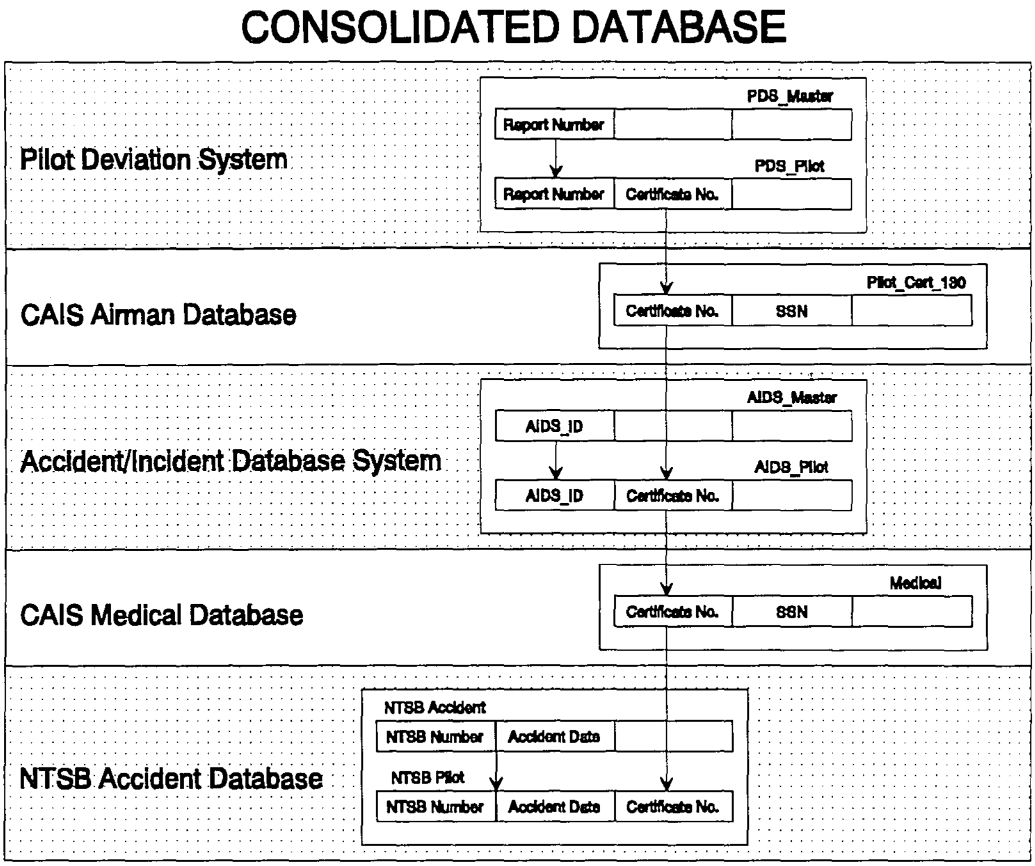

Figure 1. A view of the constituent databases of the CDB and how they may be linked using the pilot's certificate number

6.0 CDB LOADING

This section describes the procedures used to load the CDB information from the source databases and methods used to verify the contents.

6.1 Source Database Loading

The following techniques were utilized for transforming the data from the source databases into the format and content required by the CDB:

1. The SQL Loader provided by Oracle was used for PDS.
2. A custom-developed Pro*C load program was used for FAA Medical to assign certificate numbers based upon match of social security number, last name, date of birth, and gender.
3. A custom-developed Pro*C load program was used for NTSB because of the diversity of formats used for the three different time periods, i.e., 1976-1981, 1982, and 1983-1988.
4. A custom-developed COBOL program was used to extract the appropriate information from the CAIS Certification database which was supplied as COBOL Indexed Sequential binary files. A Pro*C program was developed to load this information into the CDB.
5. A custom-developed Pro*C load program was used for AIDS.

The loading programs converted valid dates that were stored in a variety of formats in the source database to Oracle standard date form, e.g., dd-MON-yy where:

dd is the number of the day in the month,

MON is a three character abbreviation of the month in upper case, and

yy is the last two digits of the year.

For example, the date February 14, 1992 would be stored as 14-FEB-92.

In addition, the source database fields that utilize the CDB coding scheme shown in Table 15 were converted. Where appropriate, the pilot age and year of accident and medical examination were calculated and inserted as columns in the associated Oracle table.

6.2 Validation

The contents of the CDB were verified, where possible, by comparing the number of records loaded from the media supplied by the database administrators. In addition, statistical sources were consulted where available. Table 18 presents statistics on the CAIS medical records included in the CDB compared to the information shown in FAA 1990 Yearbook (Table II.B, Receipts of Medical Certificate Applications by Class). The data in Table II.B is based upon receipt of medical certificates, while the data from the CDB is based upon year of medical examination and only includes those applications that were acceptably processed into the CAIS system.

The large discrepancy for 1986 data shown in Table 18 was investigated further. Table 19 presents a month by month analysis of the CDB data versus the information presented in the FAA 1986 Yearbook (Table II.B, Receipts of Medical Certificate Applications by Class). It appears that the missing data for 1986 has been irrevocably lost. The data for May, June, July, and August of 1986 show large variations between that shown in the FAA Yearbook.

Table 18. A Comparison of the Number of Medical Certificate Applications Reported in the FAA Statistical Handbook with the Number of Records Loaded in the CDB for the Years 1976-1988 (The large discrepancy for 1986 is apparently due to lost records.)

Year	FAA Statistical Handbook	CDB Database	Difference
1976	542,159	524,449	17,710
1977	550,243	538,350	11,893
1978	565,534	558,973	6,561
1979	550,188	551,905	-1,717
1980	529,051	504,015	25,036
1981	529,418	502,098	27,320
1982	463,241	479,916	-16,675
1983	477,905	490,271	-12,366
1984	463,617	468,919	-5,302
1985	479,849	474,094	5,755
1986	474,392	347,604	126,788
1987	470,208	471,223	-1,015
1988	466,326	463,749	2,577
Total (All Years)	6,562,131	6,375,566	186,565
Total (Excluding 1986)	6,087,739	6,027,962	59,777

Table 19. Count of Medical Certificates in the CDB for 1986 as a Function of Month and Class (Note the low totals for May, June, July, and August)

Month	Class I	Percent	Class II	Percent	Class III	Percent	Total	
January, 86	12,386	33%	11,273	30%	13,626	37%	37,285	
February, 86	10,642	33%	9,957	31%	11,822	36%	32,421	
March, 86	11,562	31%	11,294	30%	14,446	39%	37,302	
April, 86	11,660	30%	11,298	29%	15,773	41%	38,731	
May, 86	9,198	31%	8,244	27%	12,602	42%	30,044	
June, 86	1,358	43%	805	25%	1,024	32%	3,187	
July, 86	1,437	31%	1,516	32%	1,750	37%	4,703	
August, 86	3,518	27%	4,403	34%	5,018	39%	12,939	
September, 86	11,815	29%	11,734	29%	16,640	41%	40,189	
October, 86	12,560	31%	11,730	29%	16,383	40%	40,673	
November, 86	10,506	33%	9,144	29%	12,046	38%	31,696	
December, 86	11,308	34%	9,736	29%	12,148	37%	33,192	
CDB 1986 Total	107,950	32%	101,134	30%	133,278	39%	342,362	
FAA 1986 Yearbook	133,304	28%	146,113	31%	194,975	41%	474,392	
Difference	25,354	19%	44,979	31%	61,697	32%	132,030	28%

The information extracted from each of the source databases is stored in a series of tables in the CDB. Table 20 contains the number of records included in the CDB for each of the source databases as well as an indication of the associated time periods.

Table 20. Number of Records in the CDB based on the Source Databases

CDB Source Database	Number of CDB Records	Time Period Included in CDB
FAA Medical	6,375,566	1976-1989
NTSB Accident	47,616	1976-1988
FAA Certification	2,954,999	1962-1988
FAA Pilot Deviation System	13,895	1986-1990
FAA Accident/Incident	41,341	1986-1991
Total	9,433,417	

6.3 Data Purification

The potential usefulness of the information contained in the CDB is dependent upon the internal integrity and validity of the contents associated with each of the source databases. Numerous anomalies in the data, such as missing values, data entry errors, out-of-range values, incorrect domain values, undocumented codes, deviant spellings, and inappropriate logical types, were encountered in the data. The source databases employed a variety of mechanisms to indicate unknown or invalid information. For example, the NTSB databases populated database fields with 9's or Z's to indicate this condition. However, this variety of mechanisms made formulating queries in such a fashion to exclude unknown/invalid information very difficult. Therefore, any consistent rule that was detected to indicate that the source database information was unknown or invalid was used during the load process to convert the information to Oracle NULL fields which are ignored when retrieving information from the database. Invalid conditions that resulted in the assignment of a NULL value includes:

- A character contained in a numeric field
- Incorrect dates such as February 29, 1991, November 31, 1989, etc.

- For those values converted to CDB codes as listed in Table 15, an unrecognized code value was stored as NULL.

Flight information fields contained in the NTSB data format prior to 1982 were appended with an A to indicate pilot or a B to indicate co-pilot and were stored as character fields. These fields were stripped of the trailing character and converted to numeric values. The trailing character was used to determine whether the information should be stored for the Pilot or Co-pilot.

In some cases, invalid or incomplete information can be inferred from other elements such as if the pilot age is blank or invalid, it can sometimes be calculated using the date-of-birth element. However, these types of corrections were not made. Some database elements from the original sources contain inherent factors that tend to result in confusion. For example, the recent flight time information in CAIS is for a six month period while the information in NTSB is for a 90 day period. One technique to prevent problems when using recent flight hours in analysis would be to add an additional element to the NTSB information in the CDB that represents recent flight time over a six month period. This new value would be calculated by multiplying the 90 day information by 2. The original 90 day information would be preserved. However, based upon guidance from CAMI, the information obtained from the source databases was left intact and no attempt was made to correct the data anomalies or incorporate additional data elements.

7.0 CDB ACCESS METHODS

The following techniques are available to access the CDB to formulate a CDB data request and obtain the desired information:

1. Use the Intelligent Study Builder (ISB) on a PC,
2. Use SQL*Plus to specify an ad-hoc query, and
3. Develop a Pro*C program that performs a specialized type of data study that cannot be handled by techniques 1 and 2.

The ISB is a microcomputer-based tool for automating the specification of the desired information from the CDB. The ISB can be used in two ways:

1. To specify an ad-hoc database query and
2. To calculate accident rates as used in the series of experiments to examine various factors that impact aviation accidents.

The ISB, as described in the CDB User's Guide (Harris et al, 1992), is a microcomputer-based tool that provides a user-friendly, windows-based interface to assist users in formulating their queries. It provides a point-and-click user-computer interface that displays the list of columns associated with each table. In addition, it embeds the code tables for the encoded data columns to assist the user in determining the appropriate values. After completing their query, the user submits the resulting files to the CAMI CDB point-of-contact for processing on the HP Workstation. The user subsequently will receive a floppy diskette containing the results of their queries.

SQL*Plus is Oracle's interactive query product that provides a pseudo English-like language for formulating queries. It permits users to formulate a query using the SQL Select statement that specifies the information to be displayed as a result of the query and the criteria for selecting the desired records.

Pro*C provides the capability to embed Oracle statements within programs developed using the C language. It allows the formulation of any type of search request that can be specified using the SQL Select statement. However, the processing of the selected records is not limited to those provided by SQL*Plus and virtually any process that can be written or embedded in C can be performed. It also generates very efficient code and allows the retrieval of multiple records with a single Oracle call. The Pro*C capability is summarized in the CDB Programmer's Reference (Harris et al, 1992).

The advantages and disadvantages of each of the techniques are summarized in Table 21.

Table 21. Advantages and Disadvantages of CDB Access Methods

Feature	Intelligent Study Builder	SQL*Plus	Pro*C
Ease of Use	Easy	Medium	Difficult
Computer Knowledge	Minimum, mainly about types of information required.	Requires knowledge of Oracle and SQL*Plus.	Requires knowledge of Oracle, C, and Pro*C methods.
CDB Knowledge	Minimum, ISB provides lists of database columns and associates coding schemes.	Requires knowledge of CDB tables, columns, and relationships to other tables as well as coding schemes and interpretation.	Requires knowledge of CDB tables, columns, and relationships to other tables as well as coding schemes and interpretation.
Flexibility	Suitable for general queries.	Can handle more sophisticated queries.	Can handle extraction and processing of any procedures that can be coded using the C programming language.

Table 20. Advantages and Disadvantages of CDB Access Methods (Continued)

Feature	Intelligent Study Builder	SQL*Plus	Pro*C
Limitations	Only COUNT group function provided. No text searching capabilities provided. Column comparisons limited to values. No support for sorting results.	Computational procedures limited to those provided by SQL*Plus.	Can perform any type of calculation or procedures on extracted data.
Efficiency	Requires a pass of the appropriate database tables for each query and processes a relation at a time.	Requires a pass of the appropriate database tables for each query and processes a relation at a time.	Can perform several analyses with one pass of the appropriate database tables and can retrieve multiple rows at a time.
Access Mode	Interactive specification of query request on PC with Batch submittal of query to HP Workstation for processing	Interactive	Batch
Environment	PC front-end with HP Workstation query processor	HP Workstation	HP Workstation
Output Format	Comma delimited tabular format.	Space delimited tabular format.	Program defined.

REFERENCES

Boeing Computer Services. (1986). Aviation Standards Accident/Incident Data System (AIDS): User's Guide.

Civil Aeromedical Institute. (1981). CAIS (CAIS-AA) User's Guide. April 6, 1981.

Civil Aeromedical Institute. (1984). Aeromedical Certification Systems Manual.

Civil Aeromedical Institute. (1991). Aeromedical Certification Statistical Handbook. Report No. AC8500-1.

Codd, E. F. (1970). A relational model of data for large shared data bases. Communications of the ACM, 13, 377-87.

Codd, E. F. (1972a). Further normalization of the database relational model. In Data Base Systems (ed. R. Rustin). Courant Computer Science Symposia Series 6, Prentice Hall: Englewood Cliffs, NJ.

Codd, E. F. (1972b). Relational completeness of database sub-languages. In Data Base Systems (ed. R. Rustin). Courant Computer Science Symposia Series 6, Prentice Hall: Englewood Cliffs, NJ.

Codd, E. F. (1979). Extending the Database Relational Model to Capture More Meaning. ACM Transactions on Database Systems, 4, 397-434.

FU Associates, Ltd. (1989). Pilot Deviation System (PDS): Data Dictionary. Arlington, VA.

Golaszewski, R. (1983). The influence of total flight time, recent flight time and age on pilot accident rates. Technical Report DTRS57-83-P-80750. Bethesda, MD:Acumenics Research and Technology, Inc.

Harris, R., Hillman, D., and Voros, R. (1992a). Age 60 Consolidated Implementation. Hilton Systems Technical Report 8025-3D, Cherry Hill, NJ: Hilton Systems, Inc.

Harris, R., Hillman, D., and Voros, R.(1992b). Age 60 Consolidated Database User's Guide. Hilton Systems Technical Report 8025-3E, Cherry Hill, NJ: Hilton Systems, Inc.

Harris, R., Voros, R., and Hillman, D. (1992). Age 60 Consolidated Database Programmer's Reference. Hilton Systems Technical Report 8025-3F, Cherry Hill, NJ: Hilton Systems, Inc.

Hilton Systems, Inc. (1991). Hardware and Software Evaluation. Hilton Systems Technical Report 8025-3G, Cherry Hill, NJ: Hilton Systems, Inc.

Hyland, D.T., Kay, E. J., Deimler, J.D., & Gurman, E.B. (1992). Airline pilot age and performance: A review of the scientific literature. Technical Report 8025-1A. Cherry Hill, NJ: Hilton Systems, Inc.

Kay, E., Harris, R., Voros, R., Hillman, D., Hyland, D., and Deimler, J. (1992). Age 60 Project Consolidated Database Experiments Final Report. Hilton Systems Technical Report 8025-3C, Cherry Hill, NJ: Hilton Systems, Inc.

National Transportation Safety Board. Aviation Coding Manual. Washington, DC. NTSB.

National Transportation Safety Board. (1981). Manual of Code Classifications, Aircraft Accidents and Incidents. 4th Edition, Washington, DC. NTSB.

Oracle. (1991). SQL Language Reference Manual. Redwood Shores, CA. Oracle Corporation.

Oracle. (1991). SQL*Plus User's Guide and Reference: Version 3.0. Redwood Shores, CA. Oracle Corporation.

APPENDIX A
AGE 60 PROJECT CONSOLIDATED DATABASE QUESTIONNAIRE

Section A.1 contains the questionnaire that was distributed to potential CDB users. The tabulation of the results along with user comments are described in Section A.2. The names of the respondents were replaced with a subject number.

A.1 CDB QUESTIONNAIRE

Hilton Systems, Inc. (HSI) and Lehigh University is providing support to the FAA Civil Aeromedical Institute (CAMI) that involves developing a Consolidated Database (CDB) to support aging and pilot performance research. This CDB will incorporate information from several existing sources, including the FAA Consolidated Airman Information System (CAIS) containing both medical history and certification data, the Accident/Incident Database (AIDS), the Pilot Deviation System (PDS), and the NTSB Accident database. In addition to providing a comprehensive source of information, a key goal in the development of CDB is to ensure that it can be be used by researchers for a wide variety of research needs.

The purpose of this questionnaire is to gain an understanding of the research needs and interests of potential users regarding their anticipated use of the CDB. The information gleaned from the questionnaire will provide information to the database designers to assist in their design and development efforts. The designers will carefully review the responses and it may be necessary to obtain clarification and further discuss the provided information (please provide your phone number so we can contact you).

The data to be included in the CDB has been collected by separate groups for very specific purposes that are rarely the same as the objectives of secondary (i.e., research) analyses of the same data. This is why it is important for us to understand the data needs and processing requirements of your research applications and to use this information in designing and developing the CDB. Clearly, the types of applications to be supported are sufficiently different from standard types of database queries that special attention must be given to understanding user perspectives on the data.

The questionnaire consists of several sections. Background information is requested in Section I while Section II presents a brief description of the databases planned for inclusion in CDB. In Section III, we ask you to describe your research interest(s) in narrative form. The purpose of the narrative is to provide the system designers with an in-depth description of your use of the CDB and the types of research questions you would pose. The data needs described in Section III will be used for testing and fine-tuning the CDB in consultation with the researcher/user. Section IV of the questionnaire asks specific questions about various aspects of your data and processing needs. Those researchers who have previously used any of the FAA and NTSB databases are requested to describe these uses in Sections V through VIII.

Thank you for taking the time to fill out the attached questionnaire. We feel that you, as a potential user of the tool, are the best source to provide information about the data needs and processing requirements for the Consolidated Database (CDB). Your suggestions and comments will be carefully considered.

The more specific and detailed your responses, the easier it will be for us to understand the information and incorporate your suggestions into the design of the CDB. If there is not enough space for your comments, please feel free to continue on the back of the questionnaire pages or attach additional pages.

We would like to contact you if we have any questions about your responses. Please fill out the following:

Name: ________________________________

Title: ________________________________

Address: ________________________________

____

____

____

Phone: ________________________________

Fax: ________________________________

Please return the completed survey to:

Regina Harris
Hilton Systems, Inc.
Suite 202
950 North Kings Highway
Cherry Hill, NJ 08034

FAA AGE 60 PROJECT CONSOLIDATED DATABASE SURVEY QUESTIONNAIRE

Section I. Background

1. Please briefly describe your current duties.

2. How many years of experience do you have for your current duties?

Please place an X in the box marked None or place an X within the scale at the location that best describes your level of expertise.

3. What is your level of computer experience?

   ❑ None Novice | | | | | | Expert

4. What is your level of experience with database management systems?

   ❑ None Novice | | | | | | Expert

5. What is your level of experience with the Oracle database management systems?

   ❑ None Novice | | | | | | Expert

6. What is your level of experience with statistical software packages?

   ❑ None Novice | | | | | | Expert

7. What is your level of experience with SPSS statistical software?

   ❑ None Novice | | | | | | Expert

8. What type of computer(s) are you comfortable using (please check all that apply)?

Microcomputers	❑ DOS	❑ Windows	❑ Macintosh
UNIX Workstations	❑ DEC	❑ HP	❑ Other
Minicomputers	❑ DEC	❑ IBM	❑ Other
Mainframes	❑ DEC	❑ IBM	❑ Other

9. Please describe any previous use of automated database systems? What are the most useful features of the systems you have used? What are the most troublesome features of the systems you have used?

Section II Consolidated Database (CDB) Information Sources

2.1 FAA Consolidated Airman Information System (CAIS) Database

The CAMI Consolidated Airman Information System (CAIS) contains medical data for pilots as well as airman certificate information. The medical certification system contains information on airman applications for medical certification based upon physical standards prescribed in FAR, Parts 61, 65, 67, and 187. CAIS airman information is organized into the following two groups of information:

1. Pilot certification which provides information on certificates, ratings, accidents, violations, cancellation/revocations, and limitations.

2. Medical certification which provides information on each certification physical including date of birth, certification class, flight hours (civilian total and last six months), medical information (cardiovascular, vision, hearing, restrictions, and pathology codes).

1. What is the extent of your knowledge of the FAA Consolidated Airman Information System (CAIS) Database?

   ❑ None Minimal | | | | | | Extensive

2. Have you previously used the FAA CAIS database for any of your research?

   ❑ Yes (Please fill out Section V about your uses of the FAA CAIS database.)

   ❑ No (Go to Question 3)

3. If you have not used the FAA CAIS database for any of your research, please describe any reasons along with any potential uses that you envision.

2.2 NTSB Accident Database

The NTSB Accident Records Database contains information derived from accidents investigated by the NTSB. From 1983 to the present, data have been collected on revised form 6120.4, NTSB Factual Report Aviation. The computerized database has a very large number of database elements representing items from an inspector's checklist. Information contained in the NTSB Accident Database includes the following: pilot information (age, flight hours for specific aircraft and total — last 24 hours, 30 days, 90 days, and total), weather, cabin crew and passenger information, aircraft information, operational phase, investigation results, and causes and factors.

1. What is the extent of your knowledge of the NTSB Accident Database?

❑ None Minimal | | | | | | Extensive

2. Have you previously used the NTSB Accident database for any of your research?

❑ Yes (Please fill out Section VI about your uses of the NTSB Accident Database.)

❑ No (Go to Question 3)

3. If you have not used the NTSB Accident database for any of your research, please describe any reasons along with any potential uses that you envision.

2.3 Accident/Incident (AIDS) Database

The Accident/Incident Database is maintained by FAA's Aviation Standards and contains data pertaining to: accident/incident location and time, pilot information including certification code and age, flight hours (total and last 90 days for all aircraft types and specific to aircraft make and model), aircraft data, investigation and remedial actions, fatalities and injuries, causes and factors, and weather conditions.

1. What is the extent of your knowledge of the Accident/Incident (AIDS) Database?

❑ None Minimal | | | | | | Extensive

2. Have you previously used the AIDS database for any of your research?

❑ Yes (Please fill out Section VII about your uses of the AIDS database.)

❑ No (Go to Question 3)

3. If you have not used the AIDS database for any of your research, please describe any reasons along with any potential uses that you envision. (After completing, go to the next section.)

2.4 Pilot Deviation System (PDS) Database

The Pilot Deviation System (PDS) database is maintained by the Office of Safety Analysis of the FAA. It contains detailed information on reported abnormal flight incidents. Information of interest includes: accident/incident description, aircraft involved, operation phase, weather conditions, flight plan, investigation results, pilot information (date of birth, all and make/model specific flight hours — total and last 90 days.

1. What is the extent of your knowledge of the Pilot Deviation System (PDS) Database?

   ❑ None   Minimal | | | | | | Extensive

2. Have you previously used the PDS database for any of your research?

   ❑ Yes (Please fill out Section VIII about your uses of the PDS database.)

   ❑ No (Go to Question 3)

3. If you have not used the PDS database for any of your research, please describe any reasons along with any potential uses that you envision. (After completing, go to the next section.)

Section III Research Interests

1. Briefly describe the applications for which you would like to use the CDB. Explain the basic framework for the hypothetical situation and/or the actual study you plan to conduct or have conducted. A profile for a hypothetical researcher is presented on the next page.

2. What do you see as the most critical features that are needed in a consolidated database system to support your research?

3. Please describe and provide point-of-contact information for any other databases that you would recommend for inclusion in CDB.

SAMPLE RESEARCH PROFILE

Part 121 pilots are subject to a mandatory age 60 retirement rule. A major problem facing the airline industry is the lack of a viable means of assessing the overall attrition rates of Part 121 pilots. At present, historical data from FAA and NTSB sources can be used to estimate the probabilities of pilots leaving the Part 121 pool (through medical disqualifications, accidents, voluntary retirements, etc.) However, this information cannot be directly used to determine when individual pilots will leave the pool. Commercial carriers are concerned about a potential problem of having an inadequate supply of experienced Part 121 pilots available, particularly determining when peak demand periods will occur for replacing pilots.

Several different methods of estimating pilot attrition rates can be contemplated, providing a range of projections for planning purposes. Each of the projections is governed by a set of assumptions, which can be varied as circumstances dictate, thus providing a "what if" capability to FAA and airline decision-makers. Some assumptions concern the recent and total flight hours of Part 121 pilots, whereby increased exposure is held to decrease attrition rates on the strength of diminished accident likelihood. Other assumptions concern the probability that Part 121 pilots will be removed from the pool before age 60. The latter types of hypotheses must be quantified by assigning numerical probabilities to various stages of age 60 retirement attainment for each pilot. That is, the probabilities of a pilot remaining in the pool for exactly 1 year, 2, years,...., n years must be estimated, where n is the anticipated span of flying until age 60 is reached. The probabilities must sum to unity.

The research effort involves establishing the empirical frequency distributions of flying hours and types of flying for Part 121 pilots. The Consolidated Database should be able to determine the initial population at time $t = 0$ of each cross-section of Part 121 pilots. Each cross-section can be thought of as generating a line of replacements or descendants, and at any given time there is a certain probability that a replacement is required for a pilot who has been removed from the pool. A pilot replacement will be one of two kinds. First, the pilot to be replaced may be of age k at time t, and is thus of age $k + n$ when replaced at time n. Second, the pilot to be replaced may have begun Part 121 service at time j ($1 <= j <= n$), so that at time n his age is $n-j$. The branch of probability theory known as renewal theory can now be used to compute the total expected number of replacements of both kinds of pilots. An advantage of this approach is that it provides a comprehensive picture of the age distributions of all Part 121 pilots over a determinate time interval. Individual pilots can be tracked as they attain age 60 and attrition rates can be computed as needed.

The project requires interfacing the Consolidated Database with a statistical package and a programming environment in which routines can be written that are not currently supported in a statistical package (such as renewal theory programs). The programming environment could be C, FORTRAN, or any other suitable scientific language.

Section IV Data and Processing Needs

1. Index of Database Elements

The CDB will contain a very large number of data elements. One technique that can be used to allow the researcher to identify those fields of interest is an index of database elements as illustrated below.

Item	Database Elements	Source
Flight hours, recent	ACC_LAST_24HOURS_ALLAC	NTSB
	ACC_LAST_24HOURS_MAKE	NTSB
	ACC_LAST_30DAYS_ALLAC	NTSB
	ACC_LAST_30DAYS_MAKE	NTSB
	ACC_LAST_90DAYS_ALLAC	NTSB
	ACC_LAST_90DAYS_MAKE	NTSB
	MED_FH_CIV_LAST_6MONTHS	CAIS
	MED_FH_MIL_LAST_6MONTHS	CAIS
Flight hours, total	ACC_TOTAL_ALLAC	NTSB
	ACC_TOTAL_GLIDER	NTSB
	ACC_TOTAL_LIGHT_AIR	NTSB
	ACC_TOTAL_MAKE	NTSB
	ACC_TOTAL_MULTI_ENG	NTSB
	ACC_TOTAL_NIGHT	NTSB
	ACC_TOTAL_OTHER	NTSB
	ACC_TOTAL_ROTORCRAFT	NTSB
	ACC_TOTAL_SIMULATOR	NTSB
	ACC_TOTAL_SINGLE_ENG	NTSB
	MED_FH_CIV_TOTAL	CAIS
	MED_FH_MIL_TOTAL	CAIS
Profession	ACC_PROFESSION	NTSB
	MED_OCCUP_CODE	CAIS

Would a listing of all database elements in the form of an index be useful?

☐ Not useful ☐ Somewhat useful ☐ Very useful

If Yes, what form should this listing of all database elements take?

☐ On-line ☐ Hard-copy ☐ Both

☐ Other (please describe)

Please provide any comments about the index of database elements example.

2. Glossary

Another technique to assist researchers in identifying the database elements of interest is a glossary. In addition to element names, a glossary can also provide descriptive information as illustrated in Figure 1 on the next page.

Would a glossary of database elements be useful?

- ❑ Not useful
- ❑ Somewhat useful
- ❑ Very useful

If Yes, what form should this glossary take?

- ❑ On-line
- ❑ Hard-copy
- ❑ Both
- ❑ Other (please describe)

If Yes, what information should be included in this glossary?

- ❑ Name of term
- ❑ Brief Description
- ❑ Coding scheme
- ❑ Associated terms
- ❑ Database source(s)
- ❑ Limitations
- ❑ Validation
- ❑ Other (Please describe)

Please provide any comments about the glossary example.

3. Data Integrity

It is anticipated that the data sources being included in the CDB will contain anomalies such as: data entry errors, out-of-range values, logical type errors, undocumented codes, computed/sanitized values, inconsistencies, etc.

Figure 1. Glossary Example[1]

Element Name	Description	Coding Scheme	Database Source	See also	Comments	Validation
ACC_ PROFESSION	Pilot profession	AM = Aircraft mechanic; D = Doctor/ dentist; F = Farmer/ rancher; PC = Pilot-civilian; PM = Pilot-military; UN=Unknown	NTSB	MED_OCCUP_CODE	Approximately 10% of entries are blank	Invalid or unknown codes are coded as UN
MED_FH_CIV _LAST_ 6MONTHS	Civilian flight hours for last six months	Number	CAIS	MED_FH_MIL_LAST_ 6MONTHS, NTS_FH_LAST_6MONTHS _ALLAC, NTSB_FH_LAST_90DAYS _ALLAC	Self-reported zero values should be treated with caution since it can represent either blank entires or actual zero flight time	Values greater than 720 hours should be treated with caution
NTSB_FH_ LAST_ 6MONTHS_ ALLAC	Civlian flight hours for last six months	Number	NTSB	NTSB_FH_LAST_90DAYS _ALLAC, MED_FH_CIV_LAST_ 6MONTHS	May want to annualize based upon medical class calculated by multiplying data for last 90 days by 2	Values greater than 720 hours should be treated with caution
MED_OCCUP _CODE	Occupation code form medical exam	AM = Aircraft mechanic, any class; C1 = Commerical or self-employed with first class; CN2 = Commercial, not self-employed with second class; CS2 = Commerical, self-employed with second class; PS = Pilot - scheduled and non-scheduled airlines; UN = Unknown	CAIS	ACC_PROFESSION	Approximately 25% of entries are blank	Invalid or unknown codes are coded as UN

[1] Coding examples do not include all current values — only illustrative examples are included.

In general, how should these anomalies be treated?

Type of Anomaly	Desired Action	
Data entry error	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous
Out of range values	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous
Logical type errors	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous
Undocumented codes	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous
Computed/sanitized values	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous
Inconsistencies	❑ Omitted from CDB ❑ Stored in separate tables ❑ Other (Please describe)	❑ Included with good data ❑ Included with good data but marked as anomalous

A frequent source of anomalies is incorrect dates such as February 29, 1991, November 31, 1989, etc. Is it acceptable to correct these dates, e.g., February 28, 1991, November 30, 1989, etc., in order to use the information in the CDB?

❑ Yes, with no markings ❑ No

❑ Yes, with marking as anomalous

❑ Other (Please describe)

In many cases, invalid or incomplete information can be inferred from other elements. For example, if the pilot age is blank or invalid, it can sometimes be calculated using the date-of-birth element. Is it acceptable to make this type of inference?

- ❑ Yes, with no markings
- ❑ No
- ❑ Yes, with marking as anomalous
- ❑ Other (Please describe)

4. Data Requirements

Is it important to indicate the source of the data elements in the CDB?

❑ Not important ❑ Somewhat important ❑ Very important

Each of the databases included in CDB has typically employed a specific coding scheme as illustrated below for medical certification database elements.

Coding Scheme Examples

Database	Element Name	Coding
CAIS	CLASS_ISSUED	11=Class 1, clear; 12=Class 1, limited; 13=Class 1, waiver; 14=Class 1, limited and waiver; 15=Class 1, special restriction and waiver; 21=Class 2, clear; 22=Class 2, limited; 23=Class 2, waiver; 24=Class 2, limited and waiver; 25=Class 2, special restriction and waiver; 31=Class 3, clear; 32=Class 3, limited; 33=Class 3, waiver; 34=Class 3, limited and waiver; 35=Class 3, special restriction and waiver; 40=AME Deferred, Case pending; 90=AME Denial

Coding Scheme Examples (Continued)

Database	Element Name	Coding
NTSB (prior to 1982)	MEDICAL_ CERTIF_PILOT	A=Class I-No waiver/limitations; B=Class II-No waiver/limitations; C=Class III-No waiver/limitations; D=Class I-With waiver/limitations; E=Class II-With waiver/limitations; F=Class III-With waiver/limitations; G=Class I-Expired; H=Class II-Expired; I=Class III-Expired; J=None; K=Foreign; Z=Unknown/Not reported
NTSB (1982)	MED_CERTIF_ CLASS_COP	I=Class 1; II=Class 2; III=Class 3
	MED_CERTIF_ COP	1=Valid - no waiver/limitations; 2=Valid - with waiver/limitations; 3=Non valid medical for this flight; 4=No medical certification
NTSB (1983 to present)	MED_CERTIF	1=None; 2=Class 1; 3=Class 2; 4=Class 3
	MED_CERTIF_ VALID	1=Valid - no waiver/limitations; 2=Valid - with waiver/limitations; 3=Non valid medical for this flight; 4=Expired; 5=No medical certification

Do you prefer numeric or alphabetic coding schemes?

❑ Numeric ❑ Alphabetic

❑ Other (Please describe)

Do you prefer that alphabetic codes be assigned alphabetically (e.g., A-H codes used in NTSB 1982) or, where possible, a representative mnemonic (e.g., 1=Class 1, 2=Class 3, 3=Class 3; EX=expired, etc.) ?

❑ Alphabetical ❑ Mnemonic

❑ Other (Please describe)

Do you prefer to have comprehensive schemes that cover all contingencies (e.g., the combination of class and any waiver/limitations such as employed in CAIS) or the development of separate codes that each represents a distinct subclass (e.g., the separation of medical certification and the associated validity as employed in the NTSB for 1983 to the present)?

❑ Comprehensive ❑ Distinct

❑ Other (Please describe)

In specifying your data requests, do you prefer using codes or actual values?

❑ Codes ❑ Actual values

❑ Other (Please describe)

Some database elements from the original sources contain inherent factors that tend to result in confusion and/or errors. For example, the recent flight time information in CAIS is for a six month period while the information in NTSB is for a 90 day period. One technique to prevent problems when using recent flight hours in analysis is to add an additional element to the NTSB information in CDB that represents recent flight time over a six month period. This new value would be calculated by multiplying the 90 days information by 2. The original 90 day information would be preserved. Do you think the use of calculated database elements such as described would be useful?

❑ Not useful ❑ Somewhat useful ❑ Very useful

Please provide any comments about adding information to CDB to minimize confusion and errors or suggest other techniques.

Approximately how many database records do you anticipate retrieving for your research needs?

❑ About 1,000 ❑ About 10,000 ❑ About 100,000 ❑ About 1,000,000

❑ More than 1,000,000 ❑ Other (Please describe)

What years of data do you anticipate as being of interest for your research needs (please check all that apply)?

❑ 1961-1965 ❑ 1966-1970 ❑ 1971-1975 ❑ 1976-1980

❑ 1981-1985 ❑ 1986-1990 ❑ 1990-present

5. Output Requirements

What will you need as database output (please check as many as needed)?

☐ Paper reports ☐ On-line computer file(s) ☐ Graphics ☐ Floppy diskette(s)

☐ Other (Please describe)

What type of processing will need to be performed on the output?

☐ Averages ☐ Counts ☐ Maximums ☐ Minimums

☐ Sums ☐ Variances ☐ Standard Deviations

☐ Other (Please describe)

6. Interfaces with Other Systems

SPSS will be available on the Hewlett-Packard workstation on which the CDB will be implemented at CAMI. Indicate below which other system interfaces might be desirable.

7. User Preferences

Who will perform CDB searches?

☐ Yourself ☐ Others

If yourself, will you be able to construct your own search strategies using the Oracle command language?

☐ No ☐ Yes

Generic search procedures can be developed to allow the user to specify parameters values in dialogue windows such as illustrated below.

Procedure 1: Recent and total flight time analysis using annualization rules						
Year	❑ 1981	❑ 1982	❑ 1983	❑ 1984	❑ 1985	
	❑ 1986	❑ 1987	❑ 1988	❑ 1989	❑ 1990	
Medical Class	❑ 1	❑ 2	❑ 3			
Age	❑ 21-25	❑ 26-30	❑ 31-35	❑ 36-40	❑ 41-45	
	❑ 46-50	❑ 51-55	❑ 56-60	❑ 61-65	❑ 66-70	
Recent Flight Time	Start	[ ]	End	[ ]	Increment	[ ]
Total Flight Time	Start	[ ]	End	[ ]	Increment	[ ]
Outputs	❑ Pilot Count			❑ Annualized Pilot Count		
	❑ Sum of recent flight time time			❑ Sum of recent flight		
	❑ Sum of total flight time			❑ Sum of total flight time		
	❑ Accident Count					

Would such generic search procedures that provide the ability to specify parameter values be useful?

❑ Not useful ❑ Somewhat useful ❑ Very useful

What are the desired response or turnaround times (understanding that the response time will be dependent upon the number of years included in the search, complexity of request, and amount of data retrieved, etc.)?

❑ Near real-time ❑ Less than 5 minutes ❑ Less than 15 minutes

❑ Less than 30 minutes ❑ Less than 1 hour ❑ Less than 4 hours

❑ Same day

❑ Other (Please describe)

8. <u>Training Requirements</u>

What type of CDB user training is preferred (please check all that apply)?

❑ Hands-on sessions ❑ Short course ❑ On-line tutorials

❑ Other (Please describe)

9. <u>Documentation</u>

What type of CDB documentation is preferred (please check all that apply)?

❑ Paper manual ❑ On-line manual

❑ Other (Please describe)

Section V FAA Consolidated Airman Information System (CAIS) Database

1. Please briefly describe how you used the FAA CAIS database for your research along with a list of the variables of interest.

2. What is the extent of your knowledge of the coding scheme employed in the FAA CAIS Database?

❑ None Minimal |__|__|__|__|__| Extensive

3. How did you obtain your research data from the FAA CAIS database?

❑ Computer Access ❑ Manual Request

4. How long did it take to obtain your research data?

❑ Real-time ❑ Less than 1 day ❑ Less than 1 week ❑ Less than 2 weeks

❑ Less than 1 Month ❑ More than 1 Month ❑ Other (Please describe)

5. What format was used to supply the results for your research request (please check all that apply)?

- ❑ Paper reports
- ❑ On-line computer file(s)
- ❑ Graphics
- ❑ Floppy diskette(s)
- ❑ Other (Please describe)

6. Approximately how many database records were retrieved for your research request?

- ❑ About 1,000
- ❑ About 10,000
- ❑ About 100,000
- ❑ About 1,000,000
- ❑ More than 1,000,000
- ❑ Other (Please describe)

7. What years of data were of interest for your research request (please check all that apply)?

- ❑ 1961-1965
- ❑ 1966-1970
- ❑ 1971-1975
- ❑ 1976-1980
- ❑ 1981-1985
- ❑ 1986-1990
- ❑ 1990-present

8. Please briefly describe what you liked about the process.

9. Please briefly describe any dissatisfaction with the process.

10. Please provide any other comments about your research use of the FAA CAIS database.

Section VI NTSB Accident Database

1. Please briefly describe how you used the NTSB Accident database for your research along with a list of the variables of interest.

2. What is the extent of your knowledge of the coding scheme employed in the NTSB Accident database?

❑ None Minimal | | | | | | Extensive

3. How did you obtain your research data from the NTSB Accident database?

❑ Computer Access ❑ Manual Request

4. How long did it take to obtain your research data?

❑ Real-time ❑ Less than 1 day ❑ Less than 1 week ❑ Less than 2 weeks

❑ Less than 1 Month ❑ More than 1 Month ❑ Other (Please describe)

5. What format was used to supply the results for your research request (please check all that apply)?

❑ Paper reports ❑ On-line computer file(s) ❑ Graphics ❑ Floppy diskette(s)

❑ Other (Please describe)

6. Approximately how many database records were retrieved for your research request?

❑ About 10 ❑ About 100 ❑ About 500 ❑ About 1,000

❑ More than 1,000 ❑ Other (Please describe)

7. What years of data were of interest for your research request (please check all that apply)?

❑ 1961-1965 ❑ 1966-1970 ❑ 1971-1975 ❑ 1976-1980

❑ 1981-1985 ❑ 1986-1990 ❑ 1990-present

8. Please briefly describe what you liked about the process.

9. Please briefly describe any dissatisfaction with the process.

10. Please provide any other comments about your research use of the NTSB Accident database.

Section VII Accident/Incident (AIDS) Database

1. Please briefly describe how you used the AIDS database for your research along with a list of the variables of interest.

2. What is the extent of your knowledge of the coding scheme employed in the AIDS database?

❑ None Minimal |___|___|___|___|___| Extensive

3. How did you obtain your research data from the AIDS database?

❑ Computer Access ❑ Manual Request

4. How long did it take to obtain your research data?

❑ Real-time ❑ Less than 1 day ❑ Less than 1 week ❑ Less than 2 weeks

❑ Less than 1 Month ❑ More than 1 Month ❑ Other (Please describe)

5. What format was used to supply the results for your research request (please check all that apply)?

❑ Paper reports ❑ On-line computer file(s) ❑ Graphics ❑ Floppy diskette(s)

❑ Other (Please describe)

6. Approximately how many database records were retrieved for your research request?

❑ About 10 ❑ About 100 ❑ About 500 ❑ About 1,000

❑ More than 1,000 ❑ Other (Please describe)

7. What years of data were of interest for your research request (please check all that apply)?

❑ 1961-1965 ❑ 1966-1970 ❑ 1971-1975 ❑ 1976-1980

❑ 1981-1985 ❑ 1986-1990 ❑ 1990-present

8. Please briefly describe what you liked about the process.

9. Please briefly describe any dissatisfaction with the process.

10. Please provide any other comments about your research use of the AIDS database.

Section VIII Pilot Deviation System (PDS) Database

1. Please briefly describe how you used the PDS database for your research along with a list of the variables of interest.

2. What is the extent of your knowledge of the coding scheme employed in the PDS database?

3. How did you obtain your research data from the PDS database?

❑ Computer Access ❑ Manual Request

4. How long did it take to obtain your research data?

❑ Real-time ❑ Less than 1 day ❑ Less than 1 week ❑ Less than 2 weeks

❑ Less than 1 Month ❑ More than 1 Month ❑ Other (Please describe)

5. What format was used to supply the results for your research request (please check all that apply)?

- ❑ Paper reports
- ❑ On-line computer file(s)
- ❑ Graphics
- ❑ Floppy diskette(s)
- ❑ Other (Please describe)

6. Approximately how many database records were retrieved for your research request?

- ❑ About 10
- ❑ About 100
- ❑ About 500
- ❑ About 1,000
- ❑ More than 1,000
- ❑ Other (Please describe)

7. What years of data were of interest for your research request (please check all that apply)?

- ❑ 1961-1965
- ❑ 1966-1970
- ❑ 1971-1975
- ❑ 1976-1980
- ❑ 1981-1985
- ❑ 1986-1990
- ❑ 1990-present

8. Please briefly describe what you liked about the process.

9. Please briefly describe any dissatisfaction with the process.

10. Please provide any other comments about your research use of the PDS database.

A.2 CDB QUESTIONNAIRE RESULTS

This section summarizes the results from the six respondents to the CDB Questionnaire shown in section A.1. One respondent only provided partial information.

Section I. Background

1. Please briefly describe your current duties.

Subject 1 - Research and Teaching.

Subject 2 - Supervisor aircraft accident research section, national coordinator medical accident investigation, aerospace medicine consultant to NTSB and FAA AAI, Pilot AVN-131 (part time), aeromedical examiner, senior, hacker.

Subject 3 - P.I. large epistudy - data correlation, data management, design etc. and academic researcher.

Subject 4 - Medical Director for Southwest Region.

Subject 6 - Personnel research psychologist - human factors

2. How many years of experience do you have for your current duties?

3. What is your level of computer experience?

None	Novice	*0*	*1*	*3*	*1*	*1*	Expert

4. What is your level of experience with database management systems?

None	Novice	*1*	*2*	*3*	*0*	*0*	Expert

5. What is your level of experience with the Oracle database management systems?

3	None	Novice	*3*	*0*	*0*	*0*	*0*	Expert

6. What is your level of experience with statistical software packages?

None	Novice	*0*	*3*	*1*	*2*	*0*	Expert

7. What is your level of experience with SPSS statistical software?

1	None	Novice	*2*	*1*	*0*	*2*	*0*	Expert

8. What type of computer(s) are you comfortable using (please check all that apply)?

Microcomputers	*6*	DOS	*3*	Windows	*2*	Macintosh
UNIX Workstations	*0*	DEC	*0*	HP	*1*	Other
Minicomputers	*3*	DEC	*0*	IBM	*0*	Other
Mainframes	*1*	DEC	*2*	IBM	*0*	Other

9. Please describe any previous use of automated database systems? What are the most useful features of the systems you have used? What are the most troublesome features of the systems you have used?

Subject 2 -- None/graphical user interface/user interface.

Subject 3 -- ISQL Informix - I'm very dependent on my programmer.

Subject 5 -- Relatively. Large Databases: Mental health patient services and demographics data, Air Traffic Control specialist FAA Academy performance data, Best feature of Air traffic databases was ease of matching to other additional databases through SPSS and quick easy analysis available with SPSS. Worst features - length of searches in mental health DB, awkward access, little or no statistical analytical ability.

Subject 6 -- I established and was responsible for maintaining a system that tracks the training performance of air traffic controllers. Contains about 28,000 records. The worst problem I have is that the information to be input is often inaccurate and has to be corrected before entry. Also, information from another personnel database is used to code values of variables in our system, however the translation is not a straight translation because the supporting database isn't entirely accurate. A lot of manual coding of information is required, and as a result, coding of the variables must be for manual interpretation and data entry. It gets to be pretty frustrating.

Section II Consolidated Database (CDB) Information Sources

2.1 FAA Consolidated Airman Information System (CAIS) Database

1. What is the extent of your knowledge of the FAA Consolidated Airman Information System (CAIS) Database?

	None	Minimal						Extensive
1			*1*		*2*		*1*	

2. Have you previously used the FAA CAIS database for any of your research?

   2 Yes (Please fill out Section V about your uses of the FAA CAIS database.)

   3 No (Go to Question 3)

3. If you have not used the FAA CAIS database for any of your research, please describe any reasons along with any potential uses that you envision.

Subject 3 -- Study impact age on performance, health status.

Subject 4 -- Have not developed any research interests at this time due to time constraints. I see great potential for using the database. I would like to see it used in aircraft accident investigation and in studies of sudden in flight incapacitation.

Subject 5 -- Medical - These data may prove extremely useful for longitudinal studies of the health status of pilots. Therefore it will be critical to be able to track individual pilots through the consolidated database. It will be of interest to search the database by types of pathologies for specific visual deficiencies. It will be important to be able to match individuals to both accidents. Also it will be important to be able to search by type of certificate and rating.

2.2 NTSB Accident Database

1. What is the extent of your knowledge of the NTSB Accident Database?

2 None	Minimal	2		1	1		Extensive

2. Have you previously used the NTSB Accident database for any of your research?

2 Yes (Please fill out Section VI about your uses of the NTSB Accident Database.)
4 No (Go to Question 3)

3. If you have not used the NTSB Accident database for any of your research, please describe any reasons along with any potential uses that you envision.

Subject 2 -- Insufficient at CAMI. Waiting for CDB!!

Subject 6 -- I could envision matching this information with records on operational errors (from an FAA database).

2.3 Accident/Incident (AIDS) Database

1. What is the extent of your knowledge of the Accident/Incident (AIDS) Database?

2 None	Minimal	1	1	2			Extensive

2. Have you previously used the AIDS database for any of your research?

2 Yes (Please fill out Section VII about your uses of the AIDS database.)
4 No (Go to Question 3)

3. If you have not used the AIDS database for any of your research, please describe any reasons along with any potential uses that you envision. (After completing, go to the next section.)

Subject 2 -- We have used them for the Cabin Safety Workshop and database. Requests lead to confrontation.

Subject 5 -- I would see these data used to supplement NTSB data. If possible we should cross check these two sources of accident data to determine overlap.

Subject 6 -- Could envision matching this with FAA's operational error database. Could also envision matching this with the NTSB database to identify similarities and differences.

2.4 Pilot Deviation System (PDS) Database

1. What is the extent of your knowledge of the Pilot Deviation System (PDS) Database?

5 None Minimal | *1* | | | | | Extensive

2. Have you previously used the PDS database for any of your research?

   1 Yes (Please fill out Section VIII about your uses of the PDS database.)

   5 No (Go to Question 3)

3. If you have not used the PDS database for any of your research, please describe any reasons along with any potential uses that you envision. (After completing, go to the next section.)

Subject 2 -- Poor reputation, however no direct experience. Awaiting CDB to decide.

Subject 4 -- Was not aware of its existence.

Subject 5 -- This database may provide source of age related information.

Subject 6 -- Again, match this with FAA's operational database. Also with NTSB and Accident Incident.

Section III Research Interests

1. Briefly describe the applications for which you would like to use the CDB. Explain the basic framework for the hypothetical situation and/or the actual study you plan to conduct or have conducted. A profile for a hypothetical researcher is presented on the next page.

Subject 1 -- Identify a group of pilots with certain characteristics and determine their history of accident involvement; compare accident involved pilots with others.

Subject 2 -- DUI/DWI and pilot outcomes - incident, accident, violation deviations; compare queries of NASA ASRS with CDB to contrast detail, # of reports; NASA voluntary vs. CDB - FAA/NTSB information gathering; examine taxonomy of human factions in CDB frequency, conditional probabilities; use the accident databases to provide the accident circumstances to supplement complement the medical and human performance data gathered at accidents, incidents; "The core of the national medical accident system."; medical - autopsy, pilot medical, hospital charts, expert opinions, survival factors; human performance -- CVR transcripts, witness interviews: FDR, Voice stress analysis (ATC - pilot, CUR recordings); flight attendant performance, egress.

Subject 3 -- Analyze in detail aging data. Try to track individuals over time; use for teaching methods in developmental psychology.

Subject 4 -- The database would be very helpful in aircraft accident investigation by providing possible trends. It would be helpful to investigate the accidents with respect to medication use, underlying medical problems in pilots, similarities in environmental factors, etc. The ability to track trends is critical in prevention. In the last year how many accidents have involved the use of cold remedies? Of those accidents who many involved the same drug? Was spatial disorientation involved? How many accidents showed the medication to be contributing?

Subject 6 -- I don't foresee using the database at this time. When/if I ever do, I'll formulate research questions at the time they are needed. Also, If I conduct multiple projects, I'll probably be asking different questions each time.

2. What do you see as the most critical features that are needed in a consolidated database system to support your research?

Subject 1 -- Flight Time for Part 121,135 (scheduled), 135 (non scheduled), 91 (business.), 91 (pleasure). Each pilot should give hours of each for last 6 months.

Subject 2 -- Ease of updating component databases; ease of query browsing; speed and report formatting.

Subject 3 -- Access to persons who know it. Phone consultation for my programmer.

Subject 6 -- Complete Documentation.

3. Please describe and provide point-of-contact information for any other databases that you would recommend for inclusion in CDB.

Subject 2 -- Medical and human performance database to be established in AAM-611.

Subject 3 -- Airline Data.

Subject 4 -- User friendly, visually informative, relational.

Section IV Data and Processing Needs

1. Index of Database Elements

Would a listing of all database elements in the form of an index be useful?

Not useful Somewhat useful *5* Very useful

If Yes, what form should this listing of all database elements take?

1 On-line Hard-copy *3* Both

1 Other (please describe)

Please provide any comments about the index of database elements example.

Subject 2 -- Could also indicate the number of cases containing the information.

Subject 6 -- I would prefer a data dictionary which would define the fields - I don't find these names to be very descriptive and would give all values associated with each field and their meanings. Its ok to sort variables into categories.

2. Glossary

Would such a glossary of database elements be useful?

	Not useful	Somewhat useful	*5*	Very useful

If Yes, what form should this glossary take?

1	On-line	*1*	Hard-copy	*3*	Both
1	Other (please describe)				

If Yes, what information should be included in this glossary?

4	Name of term	*4*	Brief Description	*3*	Coding scheme	*3*	Associated terms
4	Database source(s)	*4*	Limitations	*4*	Validation		
	Other (Please describe)						

Please provide any comments about the glossary example.

Subject 1 -- Good.

Subject 2 -- Use a user selectable filter to permit some > all display of categories.

Subject 3 -- Looks Good.

Subject 6 -- I would find a single document which would combine the elements of both the index and the glossary, to be of interest. If I don't know what a variable name means, it won't help to list it in an index. If I don't understand the naming conventions, it will be difficult to use the glossary as it is shown. I could see grouping variables alphabetically by category, as you have them in the index, but providing the information you have in the glossary.

3. Data Integrity

It is anticipated that the data sources being included in the CDB will contain anomalies such as: data entry errors, out-of-range values, logical type errors, undocumented codes, computed/sanitized values, inconsistencies, etc.

Subject 2 -- Pilots do not lie about their medical history.

Subject 6 -- You need to define what each of these are. What are computed/sanitized errors? Who computed them and why? How would they be sanitized? What is the difference between inconsistencies and logical errors? How would you know something is a data entry error unless it is inconsistent, out of range, or illogical?

In general, how should these anomalies be treated?

Type of Anomaly	Desired Action		
Data entry error	Omitted from CDB Stored in separate tables Other (Please describe)	 5	Included with good data Included with good data but marked as anomalous
Out of range values	Omitted from CDB Stored in separate tables Other (Please describe)	*1* *6*	Included with good data Included with good data but marked as anomalous
Logical type errors	Omitted from CDB Stored in separate tables Other (Please describe)	 *5*	Included with good data Included with good data but marked as anomalous
Undocumented codes	Omitted from CDB Stored in separate tables Other (Please describe)	 *3*	Included with good data Included with good data but marked as anomalous
Computed/sanitized values	Omitted from CDB Stored in separate tables Other (Please describe)	 *3*	Included with good data Included with good data but marked as anomalous
Inconsistencies	Omitted from CDB Stored in separate tables Other (Please describe)	 *4*	Included with good data Included with good data but marked as anomalous

A frequent source of anomalies is incorrect dates such as February 29, 1991, November 31, 1989, etc. Is it acceptable to correct these dates, e.g., February 28, 1991, November 30, 1989, etc., in order to use the information in the CDB?

1 Yes, with no markings *1* No
4 Yes, with marking as anomalous
Other (Please describe)

In many cases, invalid or incomplete information can be inferred from other elements. For example, if the pilot age is blank or invalid, it can sometimes be calculated using the date-of-birth element. Is it acceptable to make this type of inference?

2 Yes, with no markings No
3 Yes, with marking as anomalous
Other (Please describe)

4. Data Requirements

Is it important to indicate the source of the data elements in the CDB?

Not important *1* Somewhat important *5* Very important

Do you prefer numeric or alphabetic coding schemes?

2 Numeric *1* Alphabetic

Other (Please describe)

Do you prefer that alphabetic codes be assigned alphabetically (e.g., A-H codes used in NTSB 1982) or, where possible, a representative mnemonic (e.g., 1=Class 1, 2=Class 3, 3=Class 3; EX=expired, etc.) ?

Alphabetical *5* Mnemonic

Other (Please describe)

Do you prefer to have comprehensive schemes that cover all contingencies or the development of separate codes that each represents a distinct?

1 Comprehensive *3* Distinct

Other (Please describe)

In specifying your data requests, do you prefer using codes or actual values?

Codes *6* Actual values

Other (Please describe)

Do you think the use of calculated database elements such as described would be useful?

3 Not useful Somewhat useful *3* Very useful

Please provide any comments about adding information to CDB to minimize confusion and errors or suggest other techniques.

Subject 3 -- Raw fields - 1 database set. Calculate field - second set.

Subject 6 -- As long as calculated fields are clearly identified and estimations computations are very clear.

Approximately how many database records do you anticipate retrieving for your research needs?

About 1,000 2 About 10,000 About 100,000 About 1,000,000

More than 1,000,000 Other (Please describe)

What years of data do you anticipate as being of interest for your research needs (please check all that apply)?

2	1961-1965	*2*	1966-1970	2	1971-1975	*3*	1976-1980
4	1981-1985	*5*	1986-1990	*5*	1990-present		

5. <u>Output Requirements</u>

What will you need as database output (please check as many as needed)?

4	Paper reports	*5*	On-line computer file(s)	*3*	Graphics	*4*	Floppy diskette(s)
	Other (Please describe)						

What type of processing will need to be performed on the output?

3	Averages	*4*	Counts	*4*	Maximums	*4*	Minimums
4	Sums	*3*	Variances	*3*	Standard Deviations		
2	Other (Please describe)						

6. <u>Interfaces with Other Systems</u>

Indicate below which other system interfaces might be desirable.

Subject 2 -- IBM PC, Macintosh.

Subject 4 -- Not sure but what little work I have done, SPSS, I have not found it to be user friendly. Is there windows type statistical package?

7. <u>User Preferences</u>

Who will perform CDB searches?

4	Yourself	*2*	Others

If yourself, will you be able to construct your own search strategies using the Oracle command language?

2	No	*2*	Yes

Would such generic search procedures that provide the ability to specify parameter values be useful?

Not useful	*2*	Somewhat useful	*2*	Very useful

What are the desired response or turnaround times ?

2 Near real-time *2* Less than 5 minutes *2* Less than 15 minutes
Less than 30 minutes Less than 1 hour *1* Less than 4 hours
1 Same day
1 Other (Please describe)

8. Training Requirements

What type of CDB user training is preferred (please check all that apply)?

3 Hands-on sessions *2* Short course *1* On-line tutorials
1 Other (Please describe)

9. Documentation

What type of CDB documentation is preferred (please check all that apply)?

4 Paper manual *1* On-line manual
1 Other (Please describe)

Section V FAA Consolidated Airman Information System (CAIS) Database

1. Please briefly describe how you used the FAA CAIS database for your research along with a list of the variables of interest.

Subject 1 -- I supplied Certification #'s and date. They gave me age, flight times, accident, certification if denied, and path diagnoses and wanted ratings but did not get.

Subject 2 -- Used in contract research with Wright State University, health outcome study of disqualified Class I airmen. They required data of disqualified, reason, name and last address. They then mailed out a questionnaire to determine outcomes of pilots not carrying class II or III.

Subject 4 -- I have not used it yet. I have used CAIS to work airmen cases but not for statistical analysis.

2. What is the extent of your knowledge of the coding scheme employed in the FAA CAIS Database?

1 None Minimal | *1* | | *1* | | *1* | Extensive

3. How did you obtain your research data from the FAA CAIS database?

Computer Access *2* Manual Request

4. How long did it take to obtain your research data?

	Real-time		Less than 1 day		Less than 1 week		Less than 2 weeks
	Less than 1 Month	*2*	More than 1 Month		Other (Please describe)		

5. What format was used to supply the results for your research request (please check all that apply)?

1	Paper reports		On-line computer file(s)		Graphics	*2*	Floppy diskette(s)
	Other (Please describe)						

6. Approximately how many database records were retrieved for your research request?

2	About 1,000	About 10,000	About 100,000	About 1,000,000
	More than 1,000,000	Other (Please describe)		

7. What years of data were of interest for your research request (please check all that apply)?

	1961-1965		1966-1970		1971-1975		1976-1980
2	1981-1985	*1*	1986-1990		1990-present		

8. Please briefly describe what you liked about the process.

Subject 1 -- People very pleasant, later with a different request, very good turnaround when I was in a bind.

Subject 2 -- WSU did the work.

9. Please briefly describe any dissatisfaction with the process.

Subject 1 -- Took a long time. One request they could not fulfill.

Subject 2 -- There were several errors made with the processing of the search request. Time lag also a problem.

10. Please provide any other comments about your research use of the FAA CAIS database.

Subject 2 -- WSU contract revealed areas of concern in request problems. Incorrect Zips, unfiltered denial causes, etc. U of T contract researcher will benefit by both previous efforts and the availability of the CDB!!

Subject 6 -- Although this information is second hand, it may be worth passing along. Because researchers are not involved in responding to requests, and possibly because of the format of the data, research requests are difficult to obtain, never clear whether or not correct information was provided. Therefore in the consolidated DB, good docs should be available, information should be accessible easily and directly even if off line (such as seat belt information).

Section VI NTSB Accident Database

1. Please briefly describe how you used the NTSB Accident database for your research along with a list of the variables of interest.

Subject 1 -- 2 page briefs and data tapes. For all part 135 crashes. Variables about 50.

Subject 4 -- Used to access agricultural accidents.

2. What is the extent of your knowledge of the coding scheme employed in the NTSB Accident database?

2	None	Minimal			*1*		*1*	Extensive

3. How did you obtain your research data from the NTSB Accident database?

Computer Access *3* Manual Request

4. How long did it take to obtain your research data?

	Real-time		Less than 1 day		Less than 1 week	*1*	Less than 2 weeks
	Less than 1 Month	*2*	More than 1 Month		Other (Please describe)		

5. What format was used to supply the results for your research request (please check all that apply)?

3	Paper reports		On-line computer file(s)		Graphics	*2*	Floppy diskette(s)
	Other (Please describe)						

6. Approximately how many database records were retrieved for your research request?

	About 10		About 100		About 500	*2*	About 1,000
1	More than 1,000		Other (Please describe)				

7. What years of data were of interest for your research request (please check all that apply)?

	1961-1965		1966-1970		1971-1975	*1*	1976-1980
3	1981-1985	*2*	1986-1990	*1*	1990-present		

8. Please briefly describe what you liked about the process.

Subject 1 -- Stan Smith is phenomenal.

Subject 4 -- Very complete; multiple fields.

9. Please briefly describe any dissatisfaction with the process.

Subject 1 - Factors and causes hard to analyze.

Subject 4 -- Time to get information.

10. Please provide any other comments about your research use of the NTSB Accident database.

Subject 1 -- Better data needed to evaluate survival rates and crash outcome - e.g. more complete data on harness use.

Subject 2 -- Ask about costs. Ask a question regarding the paper work required to access databases: NTSB, CAIS, PDS. I am thinking about the difficulties of accessing separate bureaucracies each time for related requests. Letters, approval, backlogs, downtime costs.

Section VII Accident/Incident (AIDS) Database

1. Please briefly describe how you used the AIDS database for your research along with a list of the variables of interest.

Subject 1 -- To determine the number of accidents and incidents in cases and controls during 3 years prior to or following a date.

Subject 4 -- Looked up number of cases of hypoxia and sudden incapacitation for last 10 years.

2. What is the extent of your knowledge of the coding scheme employed in the AIDS database?

None	Minimal						Extensive
		1	1				

3. How did you obtain your research data from the AIDS database?

Computer Access *1* Manual Request

4. How long did it take to obtain your research data?

Real-time	*1*	Less than day	*1*	Less than 1 week		Less than 2 weeks
Less than 1 Month	*1*	More than 1 Month		Other (Please describe)		

5. What format was used to supply the results for your research request (please check all that apply)?

1	Paper reports	On-line computer file(s)	Graphics	Floppy diskette(s)
1	Other (Please describe)			

6. Approximately how many database records were retrieved for your research request?

	About 10	*1*	About 100	About 500		About 1,000
1	More than 1,000		Other (Please describe)			

7. What years of data were of interest for your research request (please check all that apply)?

	1961-1965		1966-1970		1971-1975	*1*	1976-1980
2	1981-1985	*2*	1986-1990	*1*	1990-present		

8. Please briefly describe what you liked about the process.

Subject 1 -- Obliging people.

Subject 4 -- Relatively good.

9. Please briefly describe any dissatisfaction with the process.

Subject 1 -- Unexplained failure to provide any record for certificate #'s beginning with 4 or 5 in the first batch. (later corrected).

Subject 4 -- Fields were sketchy; hard to decipher results.

10. Please provide any other comments about your research use of the AIDS database.

Subject 1 -- Still dealing with it.

Section VIII Pilot Deviation System (PDS) Database

1. Please briefly describe how you used the PDS database for your research along with a list of the variables of interest.

Subject 1 -- Looked at type of incident by time of day.

Subject 2 -- Deviations incidents which supplement accident analysis, also deviations not causing incident accident might be better trend information.

2. What is the extent of your knowledge of the coding scheme employed in the PDS database?

Computer Access	*1*	Manual Request	

4. How long did it take to obtain your research data?

Real-time	Less than 1 day	Less than 1 *1* week	Less than 2 weeks
Less than 1 Month	More than 1 Month	Other (Please describe)	

5. What format was used to supply the results for your research request (please check all that apply)?

1	Paper reports	On-line computer file(s)	Graphics	Floppy diskette(s)
	Other (Please describe)			

6. Approximately how many database records were retrieved for your research request?

About 10	About 100 *1*	About 500	About 1,000
More than 1,000	Other (Please describe)		

7. What years of data were of interest for your research request (please check all that apply)?

	1961-1965	1966-1970	1971-1975	1976-1980
1	1981-1985 *1*	1986-1990	1990-present	

8. Please briefly describe what you liked about the process.

Subject 1 -- FAA did the analysis and provided an easy to read tabulation.

9. Please briefly describe any dissatisfaction with the process.

10. Please provide any other comments about your research use of the PDS database.

Subject 1 -- I doubt that the data are useful because reporting so incomplete and biased.

Vita

Place of Birth:	Schenectady, New York
Date of Birth:	August 24, 1947
Mother:	Birdie Minerva Stewart
Father:	Donald Charles Stewart
Degrees:	B.A. History and Mathematics, 1969 State University College at Potsdam, New York

END

OF

TITLE