SAND97-2777C
SAND--97-2777C

Designing a Database for Performance Assessment: Lessons Learned from WIPP

CONF-980516--

Mary-Alena Martell
Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM 87185-1341

Al Schenker
LATA
2400 Louisiana Blvd. NE, Building 1, Suite 400
Albuquerque, NM 87110
November 4, 1997 Draft

RECEIVED
DEC 2 2 1997
OSTI

INTRODUCTION

The Waste Isolation Pilot Plant (WIPP) Compliance Certification Application (CCA)[1] Performance Assessment (PA) used a relational database that was originally designed only to supply the input parameters required for implementation of the PA codes. Reviewers used the database as a "*point of entry*" to audit quality assurance measures for control, traceability, and retrieveability of input information used for analysis, and output/work products. During these audits it became apparent that modifications to the architecture and scope of the database would benefit the EPA regulator and other stakeholders when reviewing the recertification application. This paper contains a discussion of the WPP PA CCA database and lessons learned for designing a database.

DISCUSSION

Background

Historically, the main driver for compiling parameters in a database was implementation of the WIPP PA codes with the database as an integrated link in the performance assessment process. The WIPP PA process is a highly-interactive, *human-computer environment*. The WIPP PA CCA database was designed as the integration link from the human data environment to the complex computer environment (Figure 1). The original drivers for compiling the WIPP PA database were: 1) to track the input values for the modelers, 2) to centrally locate the input values and 3) to provide consistency throughout the calculation.

The WIPP CCA PA database, also called the secondary database, contained the parameters derived from empirical data and other sources and in essence embodied the PA conceptual model(s) of the disposal system (Rechard, 1992).[2] A primary data base would contain the original data, i.e. hydrologic data obtained from testing, Figure 1 (Rechard, 1995)[3]. The parameters compiled in the WIPP PA CCA database were only the values that must be supplied in order

DISCLAIMER

for the computer codes to run. The database did not include code variables mapped to parameters, nor the pedigree of the data used to derive parameters.

At the architectural level all members of the user community were not pre-identified, so their needs were not designed into the database. For example, access to the database was not available to users at the desktop who needed report capabilities. As another example, during the WIPP CCA review process, the need arose to produce reports for the EPA and stakeholders with information that had not been stored in the WIPP PA CCA database. In one case a request was made to provide a report verifying that the values in the codes input files were in fact the values stored for the parameters in the database. A mapping of the variables to parameters would have made the report simple, however that capability was not designed into the system.

LESSONS LEARNED

Lessons learned from these experiences are:

1) The aim of the database development process should be to grow the knowledge processing environment infrastructure and incorporate the user needs are in the design. Build definition into the system for all users by identifying potential users and incorporating their needs into the design architecture.

2) In the regulatory environment, there is a critical need for efficient data centralization, processing, and storage to manage the data avalanche. Designing an integrated database architecture requires recognition that a huge level of activity occurs because of the disaggregated nature of the PA process.

3) The data process is a *user- and task*-centered problem and is extremely labor intensive. To manage this process it is important to understand the exact interactions between a human and the data; the human user is always close at hand, intimately involved with almost all steps of the process. It is therefore critical to understand exactly who the users are and what their needs will be. A well-designed database architecture provides a mechanism to manage effective communication across the project.

4) Emphasis should be placed on the traceability of data and the achievement of internal consistency. Consistency of values used in calculations is necessary for meaningful comparisons to be made amongst the PA calculations. Map all code variables to the database and have shared data structures among codes.

5) The PA project database must provide historical integrity and should be the *only version of the truth.* There should only be one project database and it must include date stamping for all data updates.

For those involved in developing an integrated database for regulatory compliance of PA, it is important to remember that the database development process:

- is a labor intensive process requiring human participation.

- is a continuous process of activities ranging from collection and entry of large amounts of data to the sharing and use of the data in a user-friendly environment.
- includes data cleaning, data model development, analysis, visualization, background knowledge, and output.
- requires the database interface to provide fast, easy, and consistent access to all data from everyone's desktop.
- must have a dedicated database administrator as the only one permitted to write to the database.

CONCLUSIONS:

The PA process is a highly interactive *human-computer environment.* The WIPP PA CCA method of centralizing the data with an integrated interface from the database to the codes provided an accessible path for traceability of variables both to the input files of the numerical codes and back to the original source. Mapping code variables to parameters would have improved this process. Designing the data process requires recognition that a huge level of activity occurs because of the disaggregated nature of the process. The database design is important at an architectural level, the modeling level, and the project management level.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

---

[1] US. DOE (Department of Energy). 1996. *Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant.* DOE/CAO-1996-2184. Carlsbad, NM: U.S. Department of Energy, Waste Isolation Pilot Plant, Carlsbad Area Office.

[2] Rechard, R. P. ed. 1992. *User's Reference Manual for CAMCON: Compliance Assessment Methodology Controller, Version 3.0.* SAND90-1983. Albuquerque, NM: Sandia National Laboratories.

[3] Rechard, R. P. 1995. *An Introduction to the Mechanics of Performance Assessment Using Examples of Calculations Done for the Waste Isolation Pilot Plant Between 1990 and 1992*. SAND93-1378. Albuquerque, NM: Sandia National Laboratories.

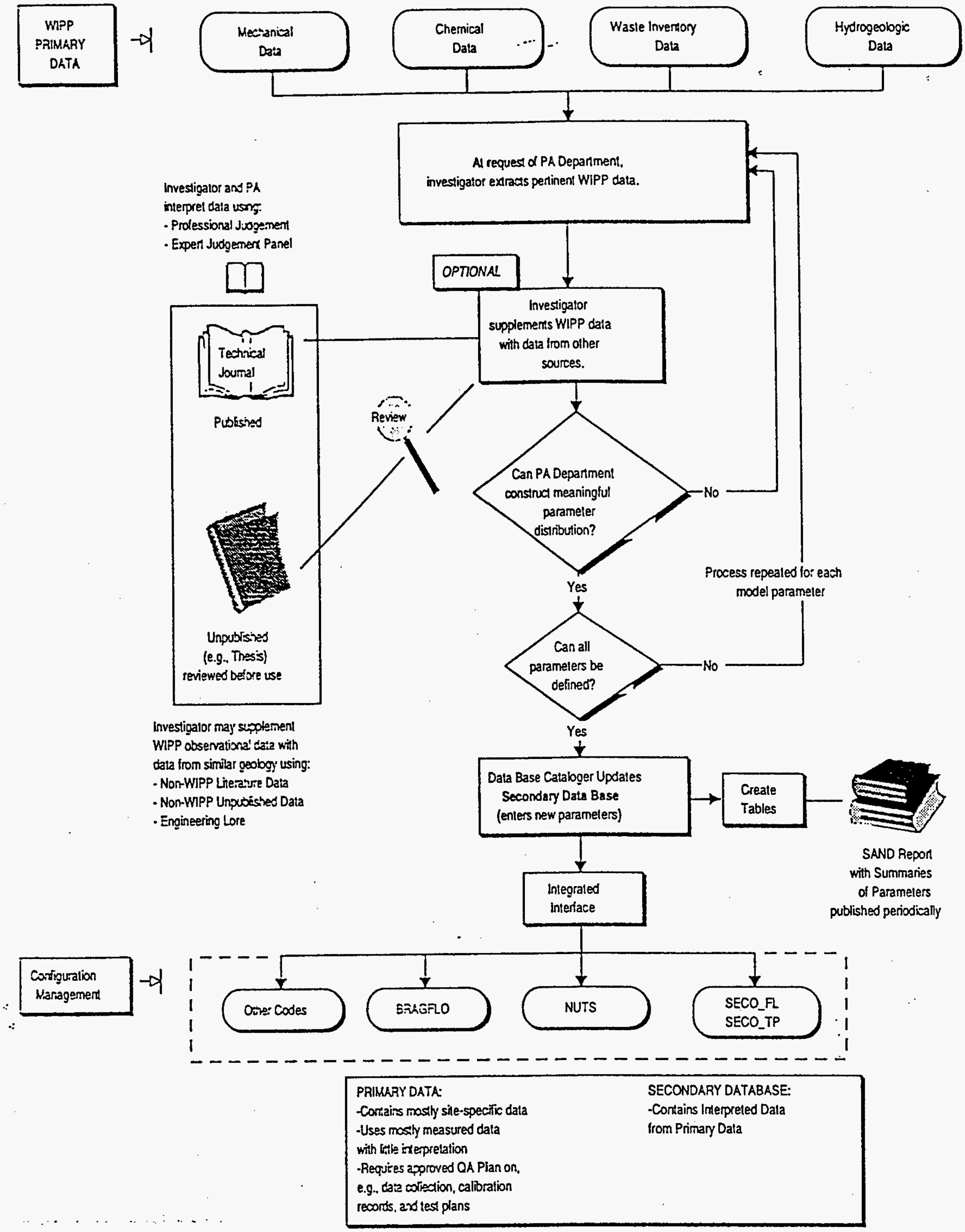

Figure 1. WIPP PA CCA database functional flow. Adapted From Rechard, 1995[2]

M98001679

Report Number (14) SAND--97-2777C
CONF-980516--

Publ. Date (11) 199712

Sponsor Code (18) DOE/EM, XF

UC Category (19) UC-2000, DOE/ER

DOE