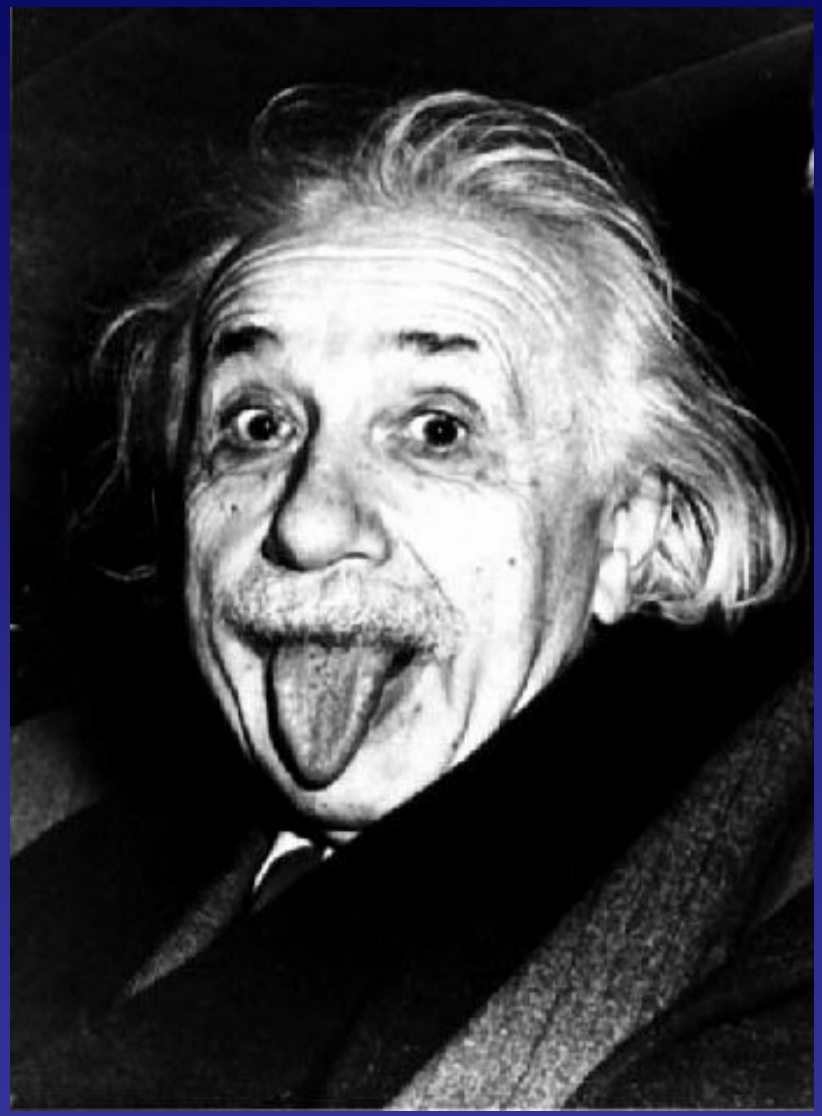
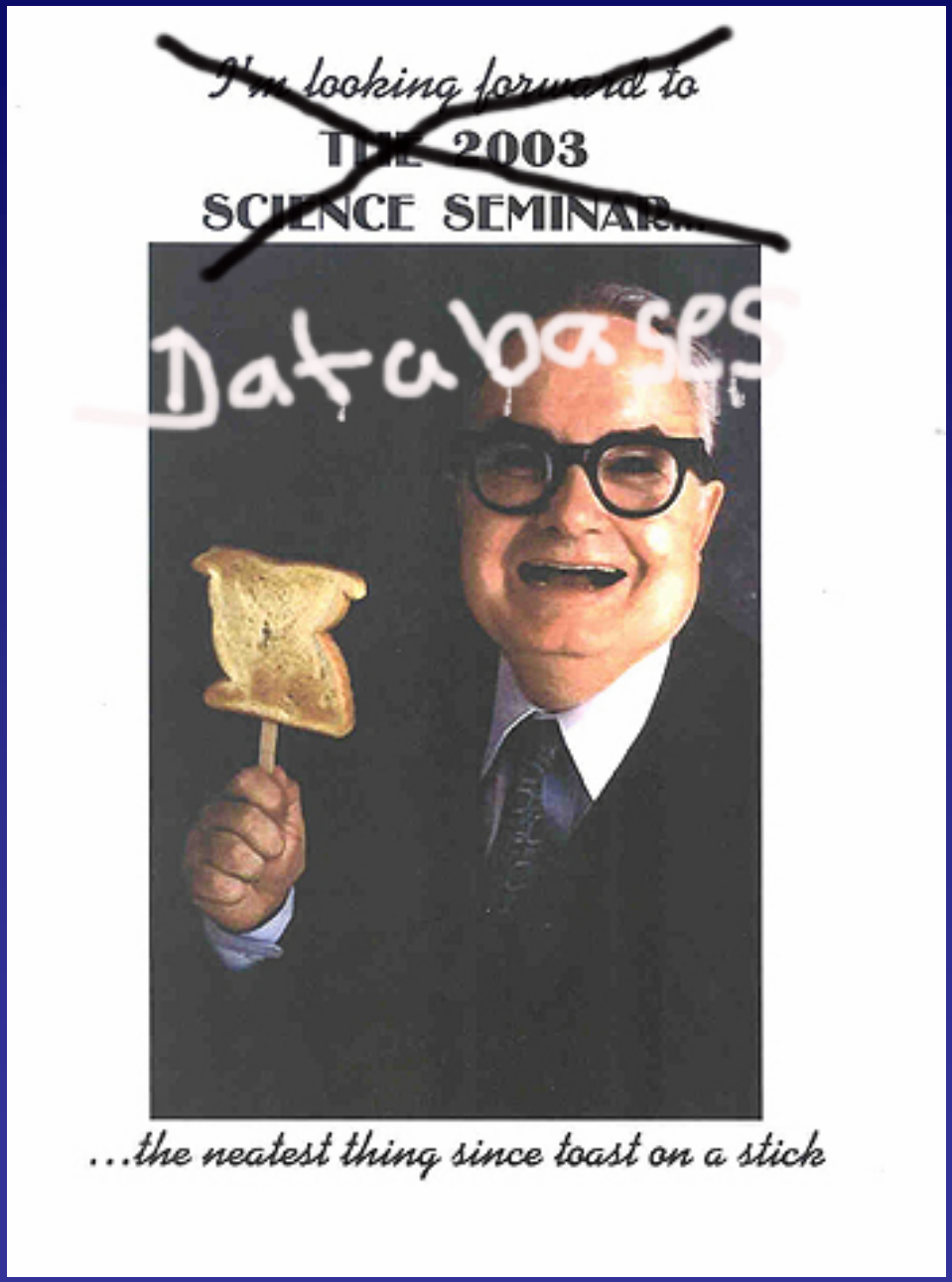
2003 WLRD Science Seminar



"A meeting of minds"


I WANT YOU
FOR U.S. ARMY
NEAREST RECRUITING STATION


I'm looking forward to
THE 2003
SCIENCE SEMINAR
Databases
...the neatest thing since toast on a stick

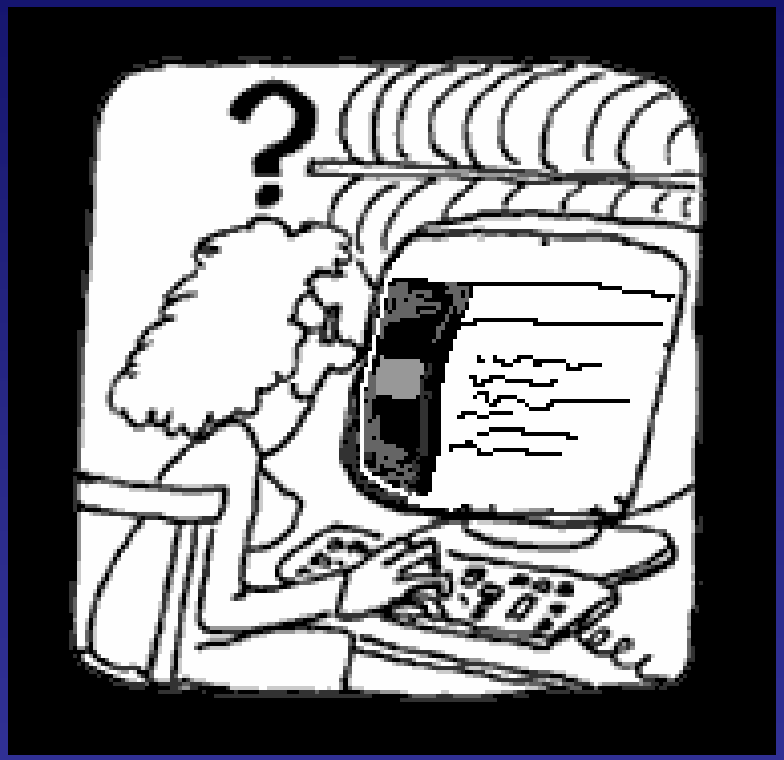

CTD Database

Introduction

- CTD data collected in Puget Sound since 1998
- 12 parameters recorded
- CTD collects at 8 hertz
- Up to 10,000 data points from a single cast
- > 11,000 Excel files

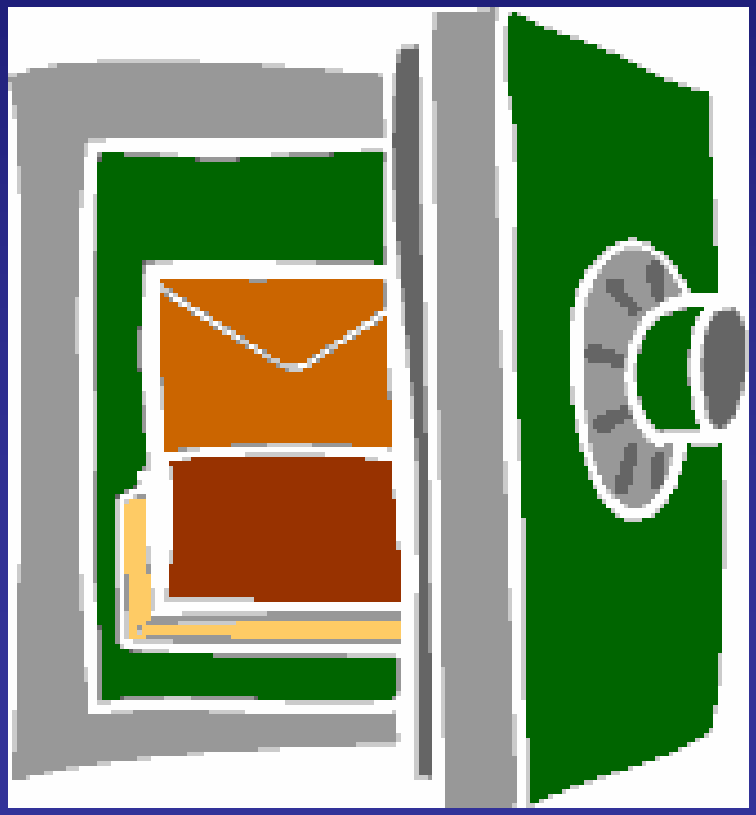
Problem Background

- LIMS stores only an abridged version of the data
- No way to attach data qualifiers
- The data cannot be queried
- Specific files difficult to find
- Data requests are time consuming
- Long-term-trend analysis of the data was difficult



Goals

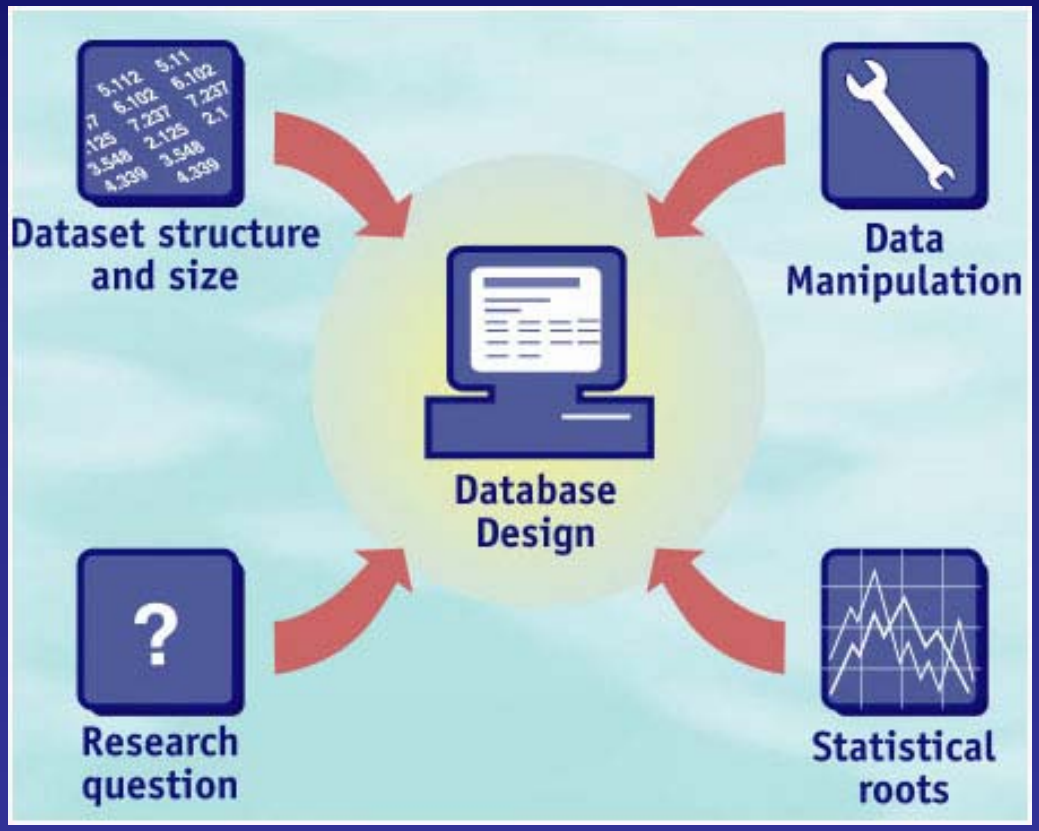
1. To build a robust, queriable relational database to securely store the current and future water quality data set.



Goals

2. To build a Web based interface that allows data users to easily download and report the data in appropriate formats.

Approach


Dataset structure and size
Data Manipulation
Database Design
Research question
Statistical roots

Approach

- Requirement specifications developed through questionnaires and interviews
- The database architecture and user interface application were designed and built to meet the needs of the users
- Meetings with a QC committee were held to determine QC policy
- Meetings with a design review committee were held throughout process to obtain feedback

Approach

- DOE model was used and modified extensively
- Data were migrated to a SQL server platform
- The user interface application was developed to optimize ease of use

Results

- **The database was built and deployed in two phases.**
- **The first phase rollout supported limited query functions. Only the 2002 water quality data was migrated to the database.**
- **A subsequent release will include a complete suite of query options and the entire data set will be migrated to the database**

Results

- **The SQL server platform has proven to be a robust, efficient and practical solution to storing a data set of this size**

- **The Web based interface is convenient and allows people to access the data without having to write SQL code**

Conclusions

- No consideration is often made for storage and retrieval of data
- Spreadsheets are inadequate
- Implementing a system for data storage and retrieval should be an integral part of any project that collects environmental data and should be addressed in the SAP
- Database architecture should be established at the project inception

Acknowledgments

- **Pava Sivam**
- **Ken Carlstedt**
- **WA DOE**
- **QC and design review committees**