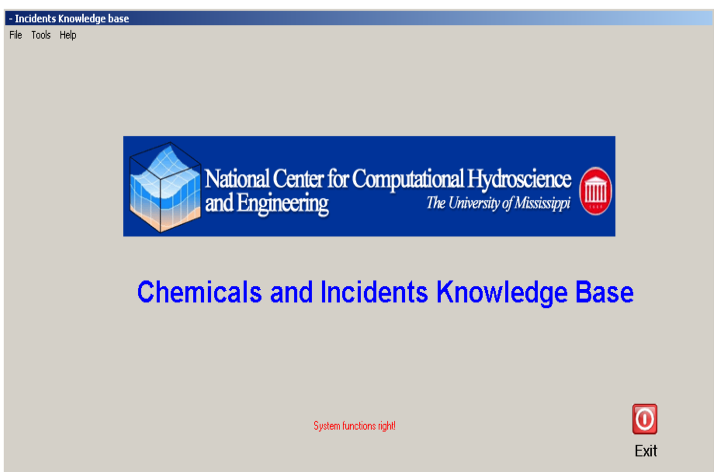
ATTACHMENT D

NATIONAL CENTER FOR COMPUTATIONAL
HYDROSCIENCE AND ENGINEERING

Development of NCCHE Chemical Spill Incident Database

Technical Report No. NCCHE-TR-2009-04 January 2009

Tingting Zhu, Yafei Jia, Xiaobo Chao, Mohamed Frihi, and Munther Hammouri



School of Engineering
The University of Mississippi
University, MS 38677

Attachment D

NATIONAL CENTER FOR COMPUTATIONAL HYDROSCIENCE AND ENGINEERING

Technical Report No. NCCHE-SERRI-TR-2009-04

Development of NCCHE Chemical Spill Incident Database

Tingting Zhu, Ph.D.
Research Scientist

Yafei Jia, Ph.D.
Research Professor

Xiaobo Chao, Ph.D.
Research Scientist

Mohamed Frihi
Research Assistant

Munther Hammouri
Research Assistant

The University of Mississippi
January 2009

Abstract

Chemical spill into water infrastructures is a serious problem threatening the general security of the people and the environment. With the development of industry, the increase of population, and the possibility of terrorist attack, the chemical spill due to accident and/or attack has become one of the main concerns of Homeland Security. The National Center for Computational Hydroscience and Engineering, the University of Mississippi, has developed a research tool: *Computational Tools for Water Security* (WIS-CSSM) sponsored by the Department of Energy, Oak Ridge Operations Office (DOE-ORO) via the Southeast Region Research Initiative (SERRI); which is an advanced computational technology for water infrastructure security [1].

The *Chemical Spill Incident knowledge Base*, entitled CHESIBASE, is a database provides information of chemical spills occurred in the past in the United States. One could review what happened at where, when, how many people were killed/injured/ hospitalized, and the amount of property damage caused, etc. The data base can be used as a reference center loaded to a computer, one could learn knowledge of chemical spills in the past and guide the practice of chemical spill emergency management. The data of CHESIBASE are loaded from the website of government agencies, and a filter was developed to remove in-significant events for emergence managements. By populating this knowledge-base with data from multiple heterogeneous databases and various source files, a reliable framework that will be used by multiple researchers is provided.

Table of Contents

Chapter 1

Introduction

The US Environmental Protection Agency and The Department of Defense are two of many agencies responsible for protecting the territory of the United States from hazards, internal and external, natural and man-made disasters. Chemical spill incident knowledge base is to be used to protect water infrastructures. Drinking water includes 160,000 public water systems which affects 84% of the nation's people and water waste that is controlled by 16,000 waste treatment plants which affects 75% of the nation's people [1]. Many of these systems are vulnerable to several types of attacks through contamination with deadly agents, the release of toxic gaseous chemicals and other means that could result in thousands of casualties as well as loss of water support for critical services. To better secure the nation's critical drinking water and waste water infrastructure, the water sector focuses on having security programs in place that enhance its ability to prevent, detect, respond to and recover from potential terrorist or other intentional acts[1].

This CHESIBASE is only a small portion of a research project for developing *Computational Tools for Water Security*, An Integrated Water Infrastructural System Chemical Spill Simulation Model (WIS-CSSM). In addition to the chemical spill incident knowledge-base, it consists of a Water Infrastructure System, Chemical property database and a scenario simulation outcome database. The computational tool is to address the needs for analysis, simulation and modeling for chemicals in the water infrastructures.

This report focuses on designing a knowledge-base framework that populates datasets from several heterogeneous databases and source files into one centralized database and provide easy reliable access to researchers within the NCCHE as well as outside organizations. The Principle Investigator of this research project is Dr. Sam, S.Y. Wang, Professor and Director of the NCCHE. This chemical incident database has been developed under direct supervision of Dr. Yafei Jia, the associate director of NCCHE; Dr. Tingting Zhu participated the supervision of the development, Mr. Mohamed Frihi developed the product.

Chapter 2

Data and Software

2.1 Data for Chemical Spill Incidents

Information on chemical spills was collected in dispersed information systems throughout government agencies. Several of the information systems within the organizations have data that is not in a unified format, spread over multiple files, formats and databases. There are several challenges that are faced when gathering information on these incidents and trying to process it:

- It is difficult to perform calculation due to difficulty of data access and the non unified schema creates misidentification of data fields and units of measure which results in a large error rate.
- It is almost impossible to execute global queries, which need records that span on multiple sources.
- Complex data queries that include multiple operations like *join* are unachievable due to the absence of consistent metadata (i.e. indices, primary key, integrity constraints)
- Very poor performance due to the extra amount of data preprocessing each time a query is executed.

In order to build a knowledge-base where detailed chemical incident information can be queried and analyzed, certain elected databases with the most relevant information had to be identified to serve as sources from which the knowledge-base is populated.

Different agencies and their resources were researched to find the most relevant chemical data. **Table 1** lists several agencies with information applicable to chemicals in the environment.

Table 1: Agencies and the details of their information systems [1]

Superfund Information Systems	The Comprehensive Environmental Response, Compensation and Liability Information System (CERCLIS) Database contains general information on sites across the nation and U.S. territories including location, contaminants and cleanup actions taken
Hazardous Substances Emergency Events Surveillance (HSEES)	Collects and analyzes information about acute releases of hazardous substances that need to be cleaned up or neutralized according to federal, state, or local law
Integrated Management Information System (IMIS)	Provided by the U.S. Department of Labor (DoL) Occupational Safety & Health Administration (OSHA) reports accidents and injuries involved.
Integrated Risk Information System (IRIS) developed EPA	It is an electronic database containing information on human health effects that may result from exposure to various chemicals in the environment
ERNS (Emergency Response Notification System) database	A database of incidents reported since 1990 to the National Response Center (NRC)
Accidental Release Information Program (ARIP) database	A collection of information on accidental releases of hazardous chemicals at fixed facilities
Information Resource System (HMIRS)	A Department of Defense (DoD) automated system developed and maintained by the Defense Logistics Agency.
Hazardous Materials Information Resource System (HMIRS)	The system assists Federal Government personnel who handle, store, transport, use, or dispose of hazardous materials. HMIRS is for official use only, unauthorized distribution of this product or its contents is strictly prohibited.

Given those initial various sources of information, the knowledge-base has to be populated from source databases that meet the requirements of the project, therefore a database is elected as a source if it provides the information needed. **Table 2** summarizes the required information.

Table 2: Criteria of information needed in knowledge-base [1]

Incidents	Date, time, location, duration, facility name and address, trasporation release, affected medium, spilled amount chemicals involved, end result/type of release, incident cause, incident short descriptions, deaths, injuries, evacuation, property damage, environment damage, stabilization/control measures, prevention/repairs, assessment of counter-measures etc
Chemical information	Chemical name, commercial name, other name, amount spiled , duration of spill, chemical particition between the water and sediment, chemical solutbility, chemical volatilization, Lethal concentration for human, fish etc, Bio-decay rate, Hydrolysis rate, Photolysis rate, Half-life, Chemcial transformation, Chemical transformation, Chemical reactivity with water, pH effect on checmical, Checmial chronicale effects on ecosystem, Bio-accumulation, maximum concentration in edible fish and shellfish etc

A total number of eight incident sources were initially found and searched for relevancy of information, but only two of them were relevant to the needs and requirements of this project, the *National Response Center (NRC) Emergency Response Notification System*, and the *Department of Health and Human Services Hazardous Substances Emergency Events Surveillance (HSEES) System*.

2.1.1 The Emergency Response Notification System (ERNS):

ERNS is a database of incidents reported since 1990 to the National Response Center (NRC). These incidents include chemical spills, accidents involving chemicals (such as fires or explosions), oil spills, transportation accidents that involve oil or chemicals, releases of radioactive materials, sightings of oil sheens on bodies of water, terrorist incidents involving chemicals, incidents where illegally dumped chemicals have been found, and drills intended to prepare responders to handle these kinds of incidents. The data includes information about who, where, when and what materials have been spilled, etc. The National Response Center is operated by the U.S. Coast Guard, and has become the central point of contact used for the reporting of many different kinds of incidents involving hazardous materials [2]

NRC data contained information on archived chemical spills from 1990 to 2006. This system also contained more than 170,818 reported incidents directly relevant to the requirements.

2 .1.2 The Hazardous Substances Emergency Events Surveillance (HSEES) System:

The HSEES system was established by the Department of Health and Human Services. “This agency is used to collect and analyze information about acute releases of hazardous substances that need to be cleaned up or neutralized according to federal, state, or local law, as well as threatened releases that result in a public health action such as an evacuation” [3]. The goal of HSEES is to reduce the morbidity (injury) and mortality (death) that result from hazardous substances events, which are experienced by first responders, employees, and the general public.

HSEES was chosen as a source database because it contained information on over 80,000 incident records based on chemical types, historical dates, chemical involved…etc

2.1.3 Limitations

The HSEES database comes in different Microsoft Excel sheets. Incidents grouped by a random number of years into one file. Each Excel file has chemical spills incidents that happened within a range of time.

The NRC database comes in 17 very large Excel workbooks representing the records of incidents that happened from 1990 to 2006, one Excel workbook file consists of multiple datasheets. Incidents of each year are put together in one file. The National Response Center publishes one file every year.

2. 2 Project Goals

In this project, the design and development of an application that, based on user configurable metadata, will be able to perform a unidirectional data mapping, by matching attributes of any database schema with our knowledge-base schema.

Once the data is aggregated from different heterogeneous sources into a consistent structured database, a second part of the project is to provide queriable data, thus maximum access to the knowledge-base, allowing large range of possible queries, and very decent performance that meets the requirements.

The challenge, beside the mapping and data migration, is to create an extensible design so that DB administrators within the NCCHE can update the knowledge-base when new information or databases become available in the future. By creating a mapping database that keeps track of metadata keys, privileged users can create and edit mapping schemas and import new datasets into the database. **Figure 1** gives an overview of the system architecture.

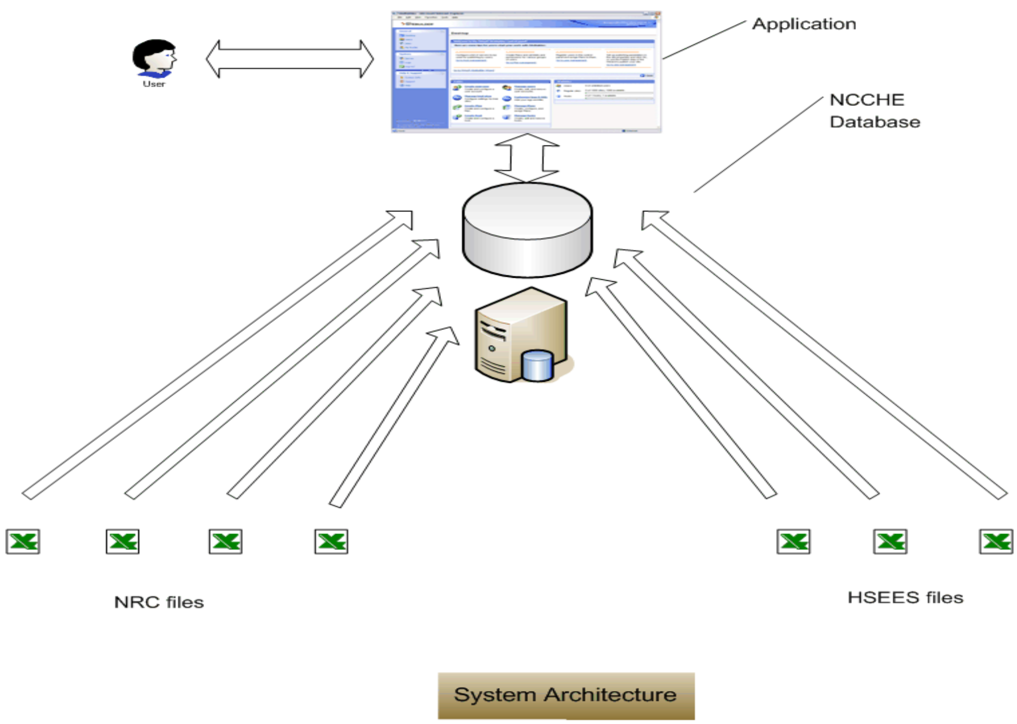


Figure1: System architecture

The main requirements of the project are:

1. Collect relevant information from HSEES and NRC databases
2. Design Knowledge-base Schema

   a. Schema must contain specific fields (such as location, chemical, damages, mitigation etc)
   b. Based on these specified fields, design schema that supports existing schemas of HSESS and NRC databases
3. Filter non-relevant information while populating knowledge-base
4. Provide a Graphical User Interface for end users
5. Create and architecture to scale for additional future records from HSEES and NRC databases
6. Create a deliverable standalone end product

2. 3 Technologies Used

Several tools and technologies have been used for designing and implementing the knowledge-base, this section provides brief description of the different tools used throughout this project.

2.3.1 MS Microsoft Access

Microsoft Access is a relational database management system (DBMS). It facilitates the storage and retrieval of structured information on a computer's hard drive. In addition to the basic DBMS features, Access provides:

- A relational database system that supports two industry standard query languages: Structured Query Language (SQL) and Query By Example (QBE).
- A full-featured procedural programming language, essentially a subset of Visual Basic.
- A simplified procedural macro language unique to Access.
- A rapid application development environment complete with visual form and report development tools.
- A various wizards and builders to make development easier [4].

Microsoft Access has been used initially as a DBMS to hold the Incidents database because if its popularity and ability to have portable databases, but later on has been replaced because of its limitations in database size and number of fields per table.

The main DBMS has been replaced by Oracle 10g.

2.3.2 Oracle XE

Oracle Database 10*g* Express Edition (Oracle Database XE) is an entry-level, small-footprint database based on the Oracle Database 10*g* Release 2 code base that is free to develop, deploy, and distribute; fast to download; and simple to administer. Oracle Database XE is a great starter database for:

- Developers working on PHP, Java, .NET, XML, and Open Source applications
- DBAs who need a free, starter database for training and deployment
- Independent Software Vendors (ISVs) and hardware vendors who want a starter database to distribute free of charge
- Educational institutions and students who need a free database for their curriculum

Oracle Database XE can be installed on any size host machine with any number of CPUs (one database per machine), but XE will store up to 4GB of user data, use up to 1GB of memory, and use one CPU on the host machine [5].

Oracle XE alleviates most of the limitations of MS Access and provides performance and easy portability therefore was chosen as the DBMS to hold the chemical incidents knowledge-base.

2.3.3 Power designer

Sybase PowerDesigner is an enterprise modeling and design solution that helps implement effective enterprise architecture. It brings powerful analysis and design techniques to the development lifecycle.

PowerDesigner combines several standard modeling techniques together with development environments, such as .NET, Java, and Eclipse, and works with more than 60 RDBMS [6].

PowerDesigner was used in this project as a supplemental tool to check the validity of data models and database schema along with aligning the user requirements with Oracle specifications. It can be interfaced with a back end Oracle RDBMS to verify the database design.

2.3.4 Programming language Borland C++

Borland C++ Builder, often abbreviated BCB, is a popular rapid application development (RAD) environment produced by the CodeGear subsidiary of Borland for writing programs in the C++ programming language.

C++ is a general-purpose, high-level programming language with low-level facilities. It supports procedural programming, data abstraction, object-oriented programming, and generic programming. Since the 1990s, C++ has been one of the most popular commercial programming languages [7].

C++ was used in this project to develop the software modules responsible for interacting directly with the source databases and destination database in the migration process as well as providing user friendly graphical interfaces and managing queries during the query process.

2.3.5 Install shield

InstallShield is a software tool for creating installers or software packages. InstallShield was also the name of the company that originally created it until it was acquired by Macrovision in 2004 [8].

Because of the complex file dependencies and configuration, Installshield was used to bundle together all the required application files, libraries and configuration settings into one software package therefore increasing portability and ease of use.

Chapter 3

Database design

Based on the nature of the information that already exists, the database design has been done in accordance to the Relational Data Model standards for database management , which mainly consists of three abstraction cycles treated separately during the project development process as shown in **Figure 2**:

- **Conceptual level :** Describes the semantics of the organization, the significant information (entities) , the characteristics of each (attributes) and the possible associations between them (relationships)
- **Logical/organizational:** Describes the data in as much details as possible, regardless of how it will be physically implemented.
- **Physical :** Describes the representation of the data that takes into account the facilities and constraints of a given Database Management System (DBMS)

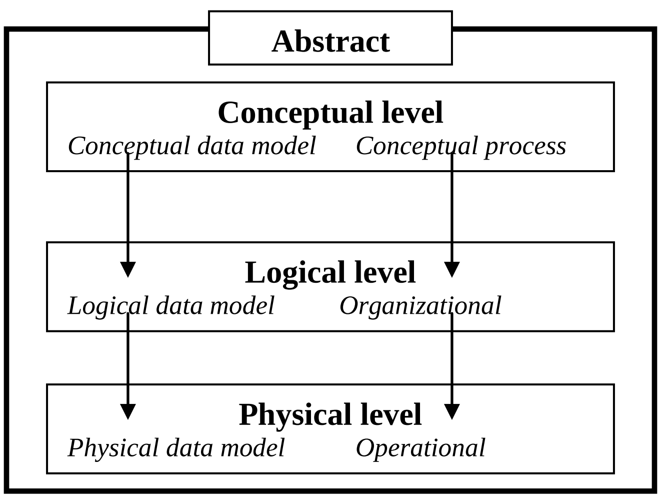


Figure 2: Abstract cycle for Relational Database Design

In order to have an efficient database design it is imperative as a first step to gather the data dictionary which consists of all the attributes of interested and then organize them in by entities based on the logical semantic relationship between the fields.

The National Response Center database has 270 attributes.

The NSEES database has 84 attributes.

The schema designed should fulfill the requirements of the incident database and at the same time support the two existing schemas as shown in **Figure 3**.

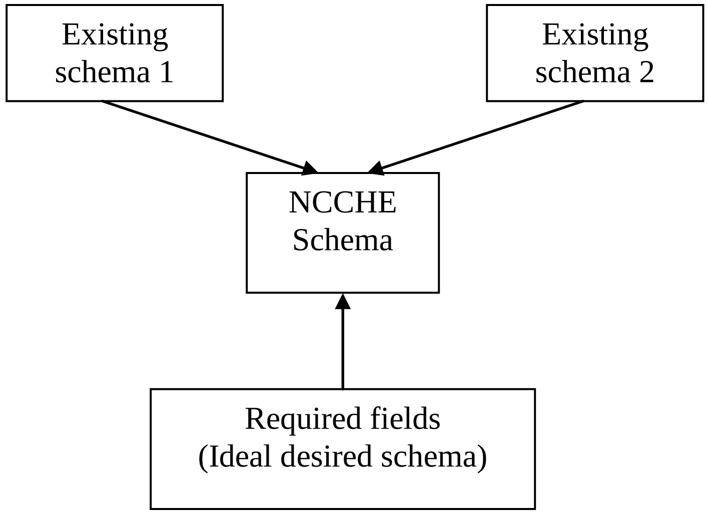


Figure 3: Schema considerations and constraints

3.1 The conceptual data model:

After removing the attributes that are not needed in this project, the new data dictionary consists of 114 relevant attributes, that are represented efficiently in the following Entity-Relationship Conceptual model conforming to the Relational Model normal forms which guarantee the ACID (Atomicity, Consistency, Integrity and Durability) model properties.

Figure 4 shows the conceptual model, where the entity “Incidents” has most of the fields organized by order of importance. The injuries were put in a separate table because in the initial source database, there were a finite constant number of injury types where each injury type is represented by a field in the incidents table and the value of that field is set to the number of people affected by that particular type of injury. Having a separate entity for injuries allows it to be more flexible towards adding new types of injuries in the future and shrinks the size of the table “Incidents” therefore speeding the query processing. The same approach has been used to move the property “chemical” into a separate entity, thus extending the relational model to support unlimited number of chemicals.

Further description of the entities and relationships is provided in the next section. See appendix 1 for the full list and description of the fields

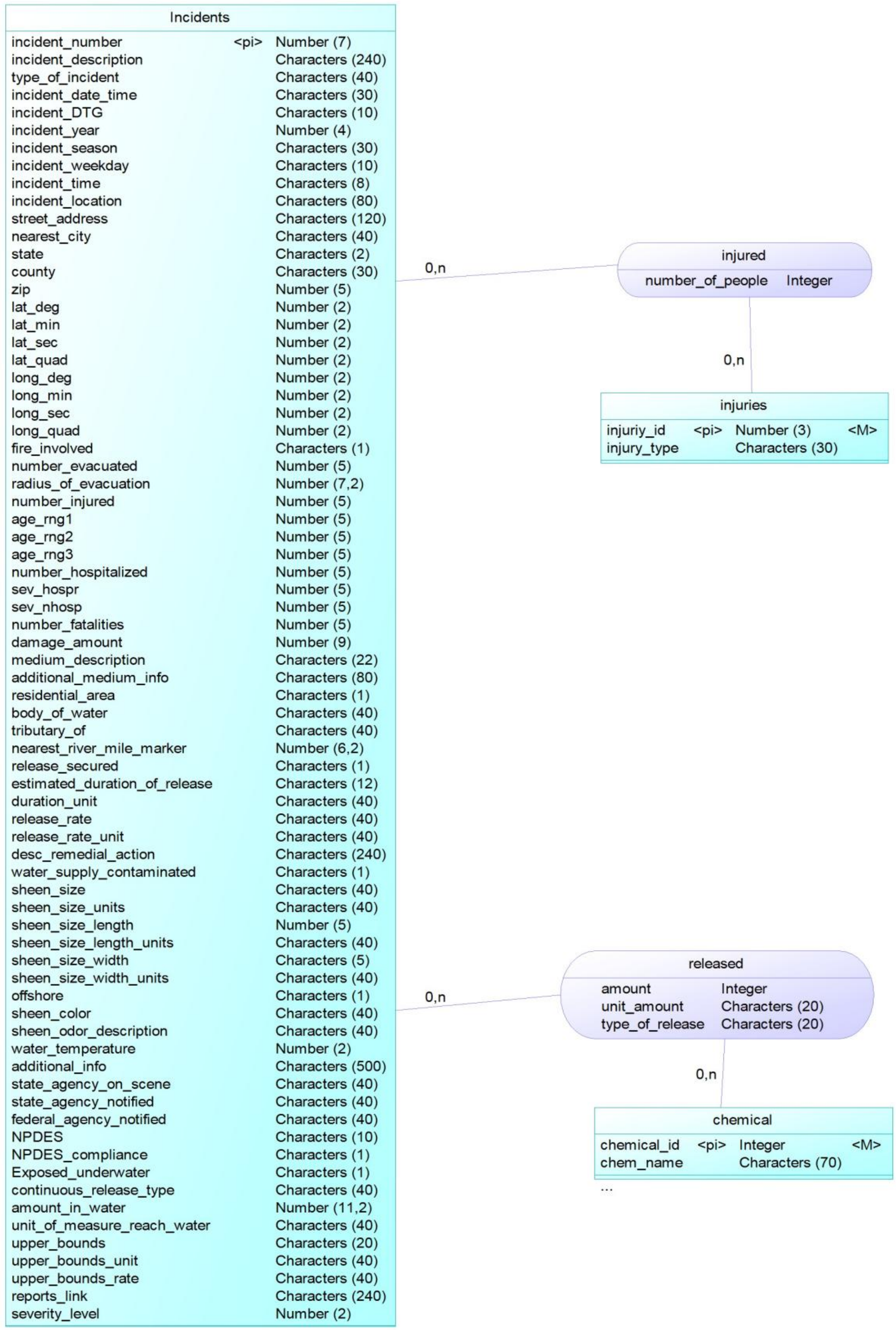


Figure 4: Conceptual Data Model

3.2 Logical Data Model

3.2.1 Entities

Incidents: Contains the attributes that are directly related to the incident such as description of the incident, location, damage…etc.

Chemical: The chemical information is presented in separate entity within the Conceptual Model as opposed to being just an attribute of the incident within the initial database tables. Placing the chemical information in a separate entity not only guarantees the third normal form for our model but also allows much better performance on chemical related queries (i.e., what are the 10 most dangerous incidents that involved Kerozene). This type of query is expected to be common, so efficiency is important.

Injuries: Even though the number of injury types was limited to 12 category of injuries according to the source database (i.e. skin injury, respiratory injury, eye injury…etc.), it is more efficient to put them in a separate table as opposed to adding 12 attribute to the incidents table, that way we eliminate redundancy and optimize queries computations.

3.2.2 Relationships

Released: Is a relationship of type many-to-many which contains 3 attributes that depend on both the incident and the chemical. A given incident may involve multiple chemicals and one chemical can be involved in multiple incidents, and for every incident-chemical pair, the relationship provides the amount spilled with the unit used to measure it and the type of release of the that chemical.

Injured: is a relationship of type many-to-many that given an incident and a certain type of injury provides the number of people affected. One incident can involve multiple types of injuries (i.e. eye injuries, skin, respiratory) and one type of injury can occur in multiple incidents.

3.3 The Physical Data Model

Taking into consideration the DBMS, the previous conceptual data model results in the following physical model under Oracle 10g DBMS constraints shown in **Figure 5**.

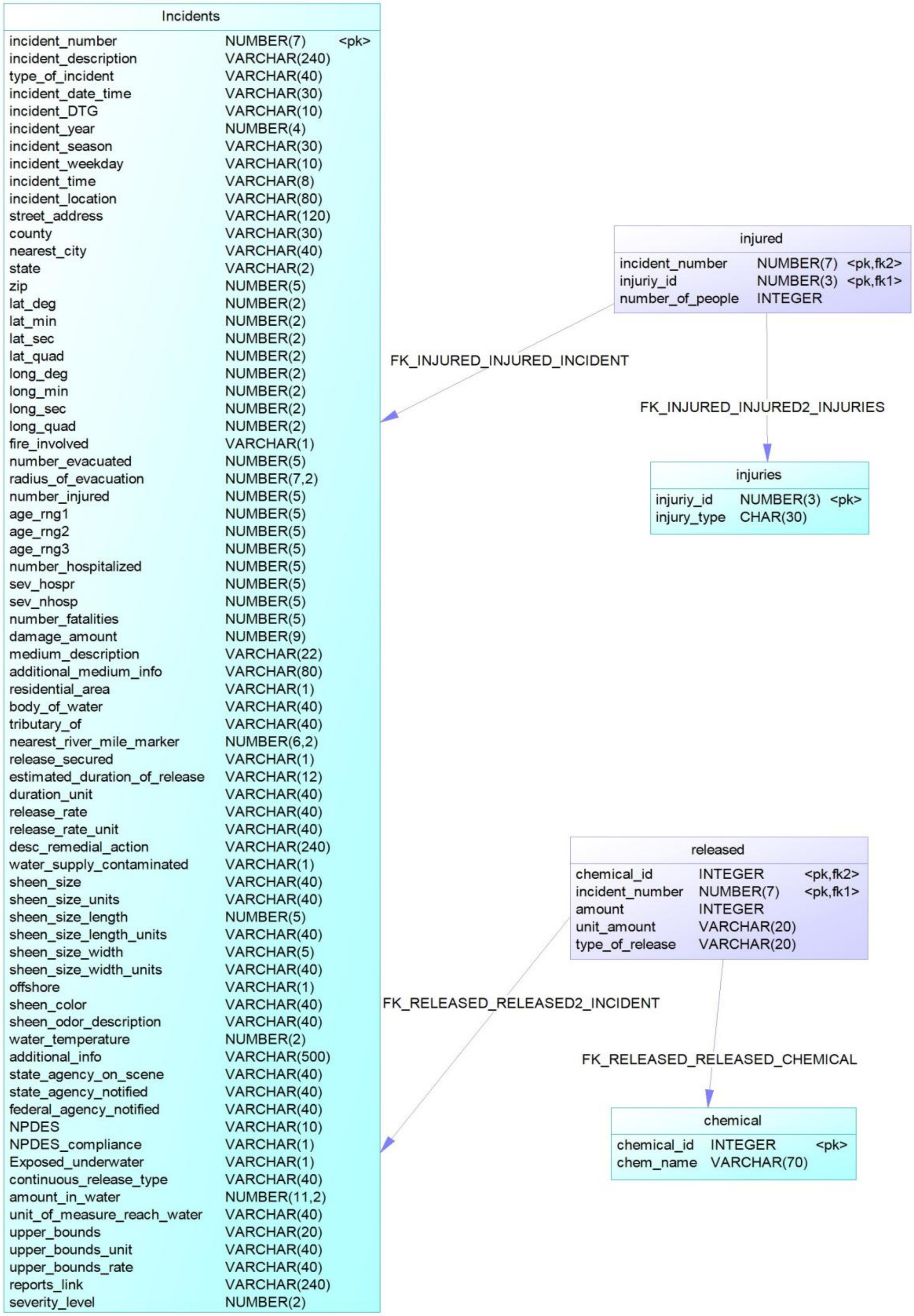


Figure 5: Physical Data Model for Oracle DBMS

This physical model translates into the following SQL Data Definition which generates the database. **Figure 6** shows a snippet of the script, for full SQL code see appendix 2.

```
/*==============================================================*/
/* Table: CHEMICAL                                              */
/*==============================================================*/
create table CHEMICAL  (
   CHEMICAL_ID          INTEGER                         not null,
   CHEM_NAME            VARCHAR(70),
   constraint PK_CHEMICAL primary key (CHEMICAL_ID)
);

/*==============================================================*/
/* Table: INCIDENTS                                             */
/*==============================================================*/
create table INCIDENTS  (
   INCIDENT_NUMBER              NUMBER(7)           not null,
   INCIDENT_DESCRIPTION         VARCHAR(240),
   TYPE_OF_INCIDENT             VARCHAR(40),
   INCIDENT_DATE_TIME           VARCHAR(30),
   INCIDENT_DTG                 VARCHAR(10),
   INCIDENT_YEAR                NUMBER(4),
   INCIDENT_SEASON              VARCHAR(30),
   INCIDENT_WEEKDAY             VARCHAR(10),
   INCIDENT_TIME                VARCHAR(8),
   INCIDENT_LOCATION            VARCHAR(80),
   STREET_ADDRESS               VARCHAR(120),
   COUNTY                       VARCHAR(30),
   NEAREST_CITY                 VARCHAR(40),
   STATE                        VARCHAR(2),
   ZIP                          NUMBER(5),
   LAT_DEG                      NUMBER(2),
   LAT_MIN                      NUMBER(2),
   LAT_SEC                      NUMBER(2),
   LAT_QUAD                     NUMBER(2),
   LONG_DEG                     NUMBER(2),
   LONG_MIN                     NUMBER(2),
   LONG_SEC                     NUMBER(2),
   LONG_QUAD                    NUMBER(2),
   FIRE_INVOLVED                VARCHAR(1),
```

Figure 6: SQL data definition code

Chapter 4

Software design

Once the Database framework is created, the next step consists of designing a user-friendly software application that allows database administrators to import, store and manage data as well as regular users to query and retrieve results from the knowledge-base. This section details the different modules of the software application and the requirements of each.

4.1 Migration Module

This software module consists of two independent sub-modules corresponding to the NRC database and HSEES database, each of them has a goal of providing the following services:

- Accepting data files from the source database
- Identifying the schema and validity of the input file
- Establish and maintain a connection to both the source database and destination database throughout the session.
- Extracting the useful records and fields from the data file
- Checking for consistency issues between the source record and existing records in the destination database.
- Update and/or append new entries to the NCCHE database based on an admin configurable mapping schema.

Figure 7 shows a flowchart of the software module in charge of the data migration.

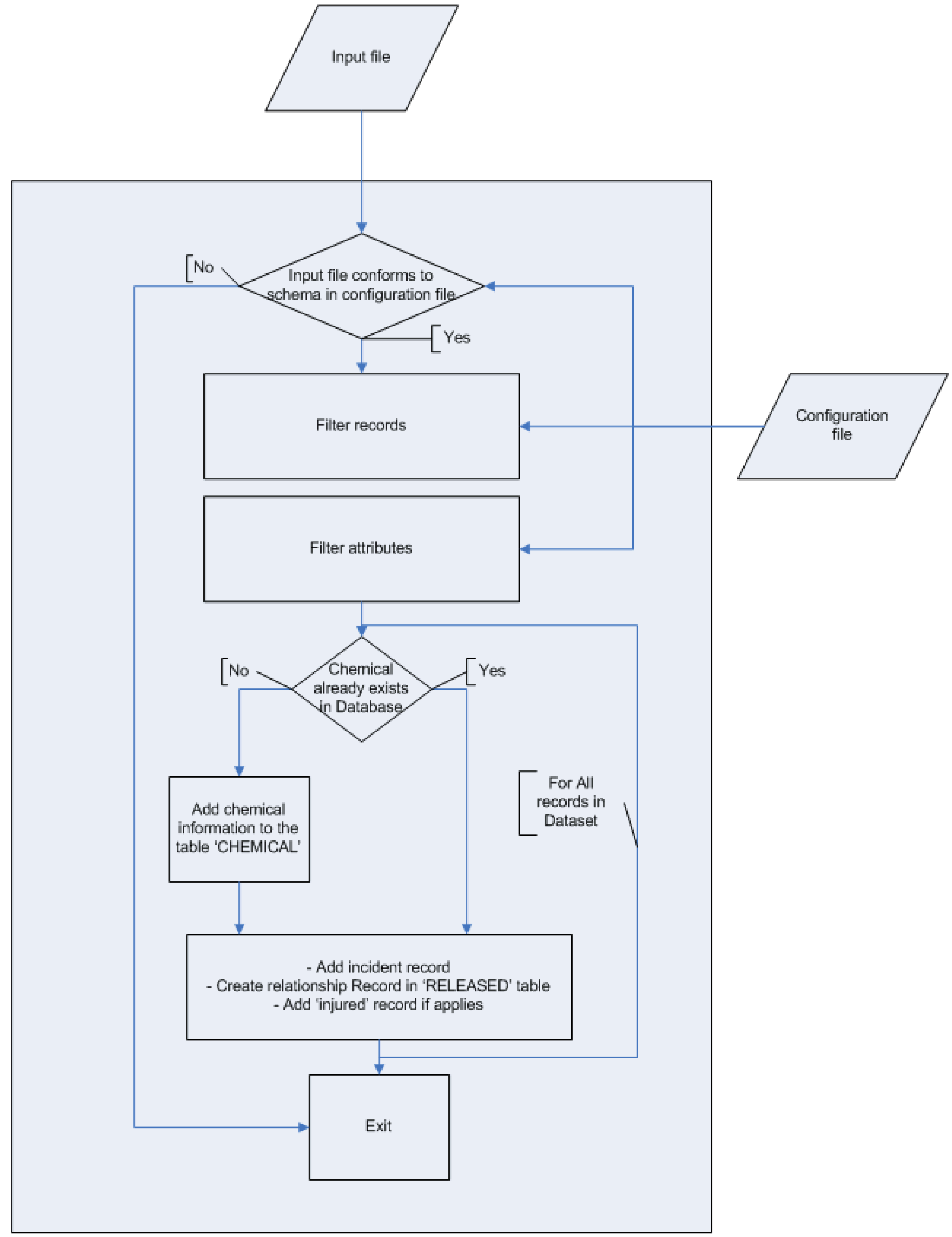


Figure 7: Migration Module flowchart

The input files are in Excel format and accessed by the application through an Object Linking and Embedding, Database (OLE-DB) component. OLE-DB is an API designed by Microsoft for accessing different types of data stored in a uniform manner, which allows developers to use Sequential Query Language capabilities to access the input data files. The migration module, after successfully connecting to the input file, proceeds by reading SQL commands from the configuration file which constitutes a mapping schema that Database administrators can later alter it to make it compatible with different database designs. **Figure 8** shows a piece of the default configuration file. These commands accessing the source files can be edited and modified to match the table name (for this example it is an Excel sheet such as “incident_commons”) and the field name (such as “seqnos” in this example which matches the incident number).

```
select

[incident_commons$].seqnos,
description_of_incident,
type_of_incident,
incident_date_time,
incident_dtg,
incident_location,
location_address,
location_nearest_city,
location_state,

…

name_of_material,
amount_of_material,
unit_of_measure,
amount_in_water,
unit_of_measure_reach_water

from
[incident_commons$],[incident_details$],[incidents$],[material_involved$]
where

[incident_commons$].seqnos = [incident_details$].seqnos and
[incident_commons$].seqnos = [incidents$].seqnos and
[incident_commons$].seqnos = [material_involved$].seqnos and
medium_desc = 'water'
```

Figure 8: SQL command extracting data from NRC input file

4.2 Data Access Module

This module is a software layer that interfaces directly with the backend populated database for the purpose of retrieving desired information based on criteria that the user specifies. This part should be designed to provide the following services:

- Interactive user-friendly interface that is easy to understand and use
- A capacity to handle a wide range of possible queries that end users might need while minimizing complexity.
- Allow users to not only select specific records but also view a selected list of attributes
- Possibility to save, load and modify user settings for a future use
- Possibility to view and manipulate the result dataset and sort it according to one of the attributes that the user chooses.
- Export the query results in a portable format for further processing. Users should be able to export, print, edit and process the results.

Figure 9 shows the flowchart diagram of the data access software module, the user chooses the search criteria though a set of selection options and the way of displaying the results then runs the query.

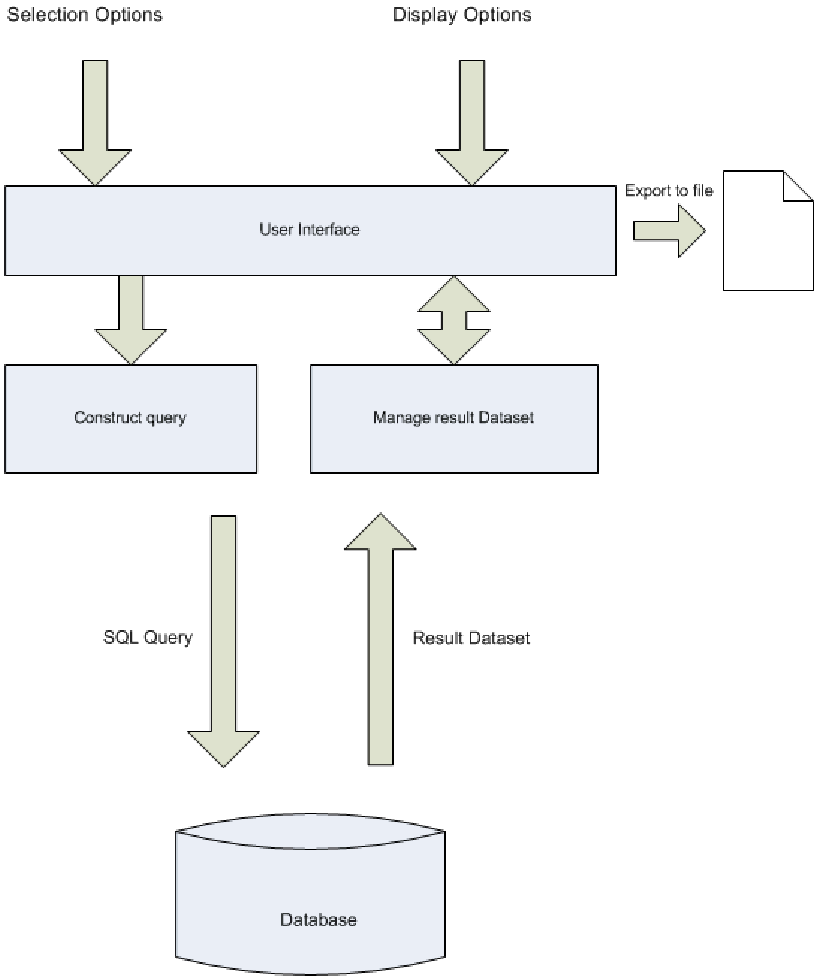


Figure 9: Data Access Module flowchart

Chapter 5

Implementation

This section describes the actual implementation of the project, the final product is packaged together to be portable and deliverable on one CD. The following screenshots explain in detail the functionality of the software.

The original CD comes with an auto-run menu showing the user a welcome screen with the installation steps as shown in **Figure 10**.

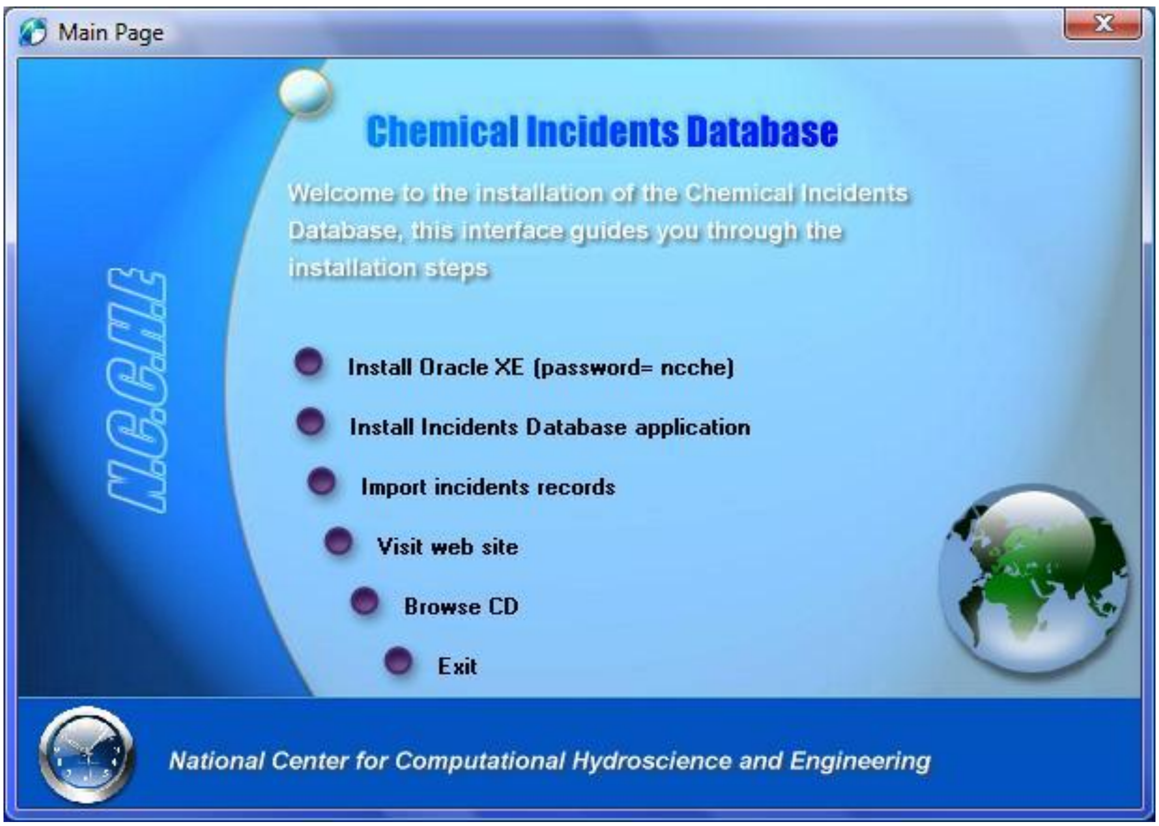


Figure 10: Installation menu

The installation package for Oracle is provided in the Original CD and the user can select the first option to install the Oracle Express Edition DBMS then follow the installation steps as shown in **Figure 11**.

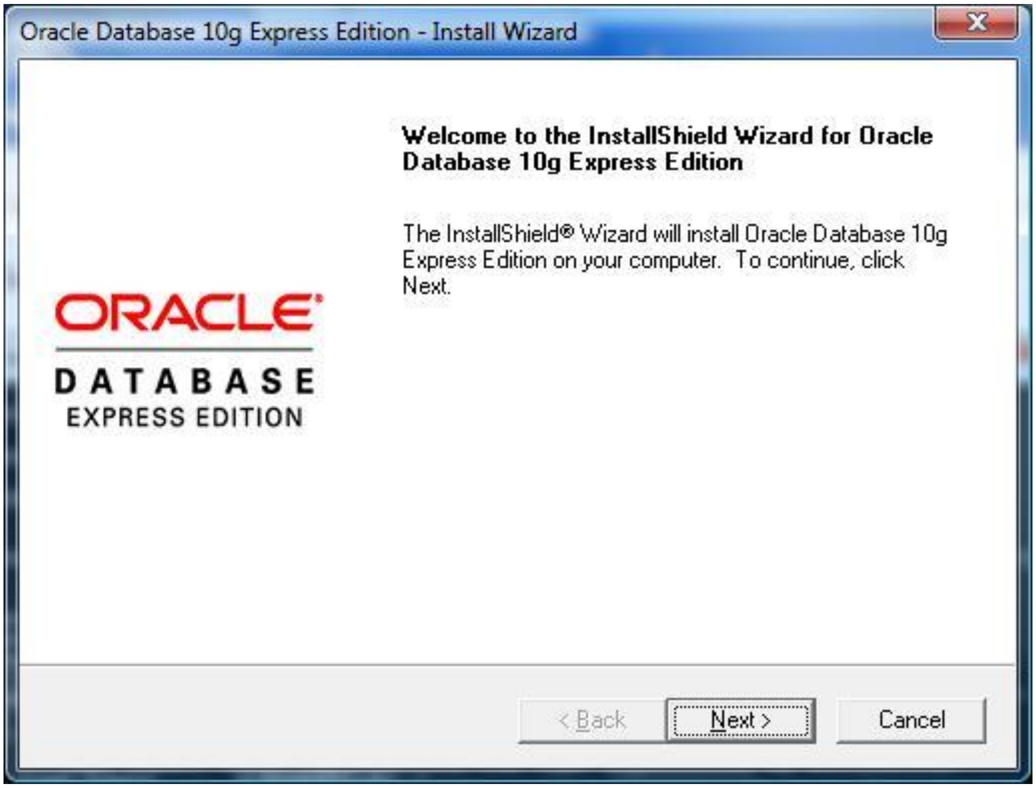


Figure 11: Oracle 10g Installation

After installing the Database management system, by clicking the option entitled "Import incidents records" the use can trigger a script which imports the schema and all the records from a backup image file of the database. **Figure 12** shows the command line window during the importation process of the records from the provided backup file.

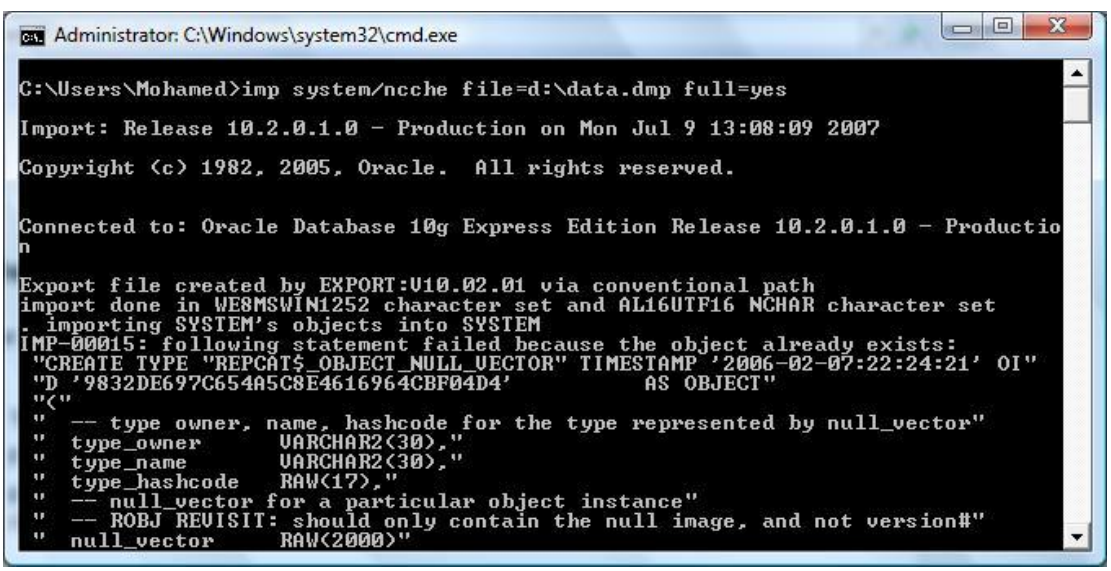


Figure 12: Importing database schema and records into Oracle

When all the data is imported, the command line window closes automatically. And gives the user the opportunity to install the application software by clicking on the item “Install incidents database application” then the welcome window in **Figure 13** appears to the user and prompt him to proceed with the installation of the software application.

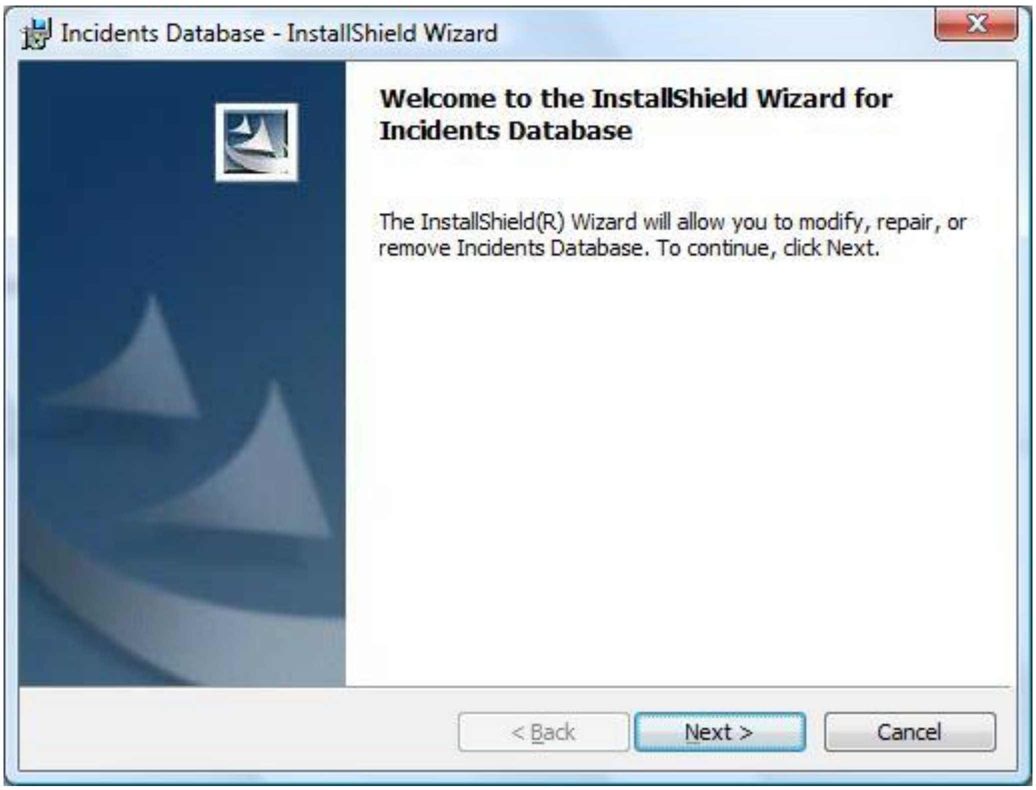


Figure 13: Incidents Database application installation

The installer creates the main executable file with additional library files, two configuration files for each of the source databases, an ODBC connection to the local oracle database and shortcuts on the desktop and start menu.

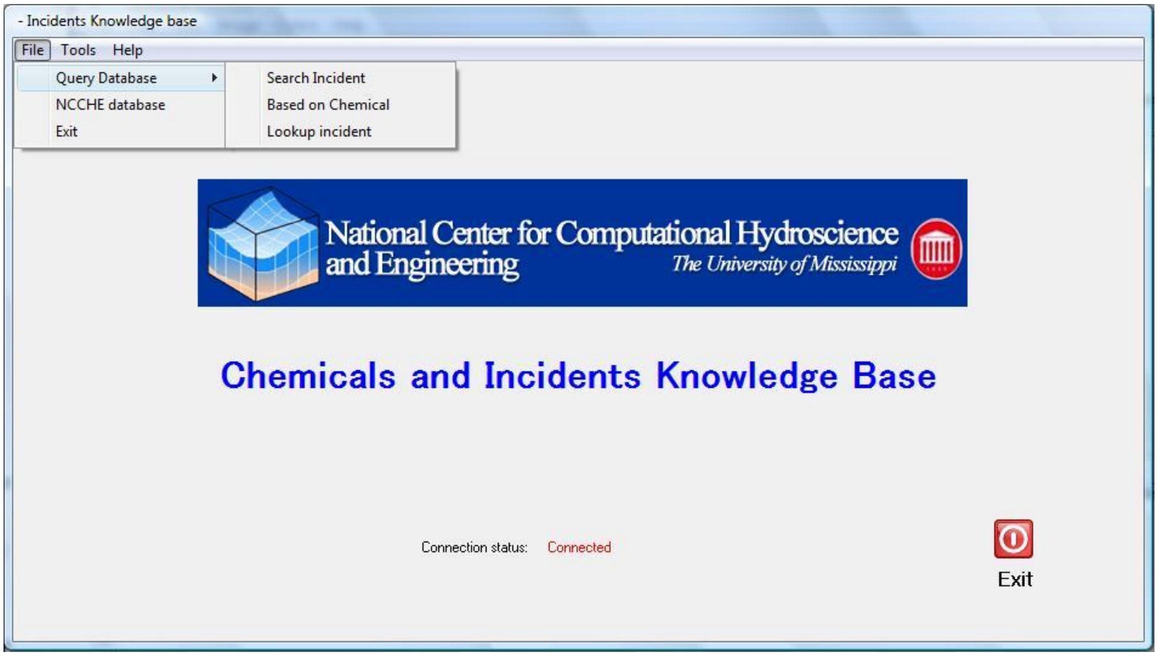


Figure 14: Application homepage

The main window shown in **Figure 14** has access to both software modules: the migration module and the data access module. By selecting from the drop-down menu, the user can choose to:

- Import new records from NRC database
- Import new records from HSEES database
- Search incidents based on multiple criteria (general search) and save results to a file
- Search an incidents based on the chemical name
- Lookup or print details about a particular incident by giving its incident code.

Figure 15 shows the "About" information under the help menu.

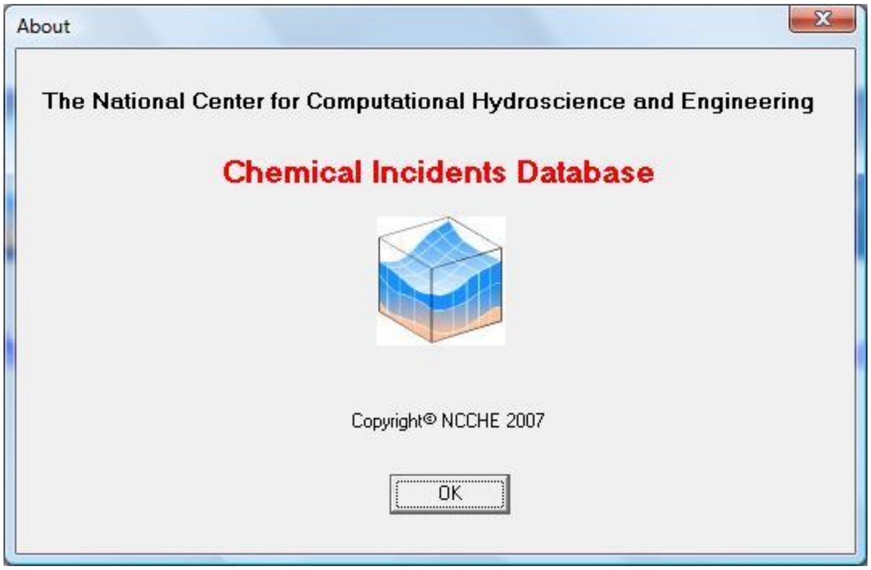


Figure 15: About information page

If the user wants to import the incidents of a particular year from the NRC database, he selects the year from a drop down box and enters the path to the input file either by typing it in the provided text box or by clicking the "Browse" button as shown in **Figure 16**, then proceeds to the next step.

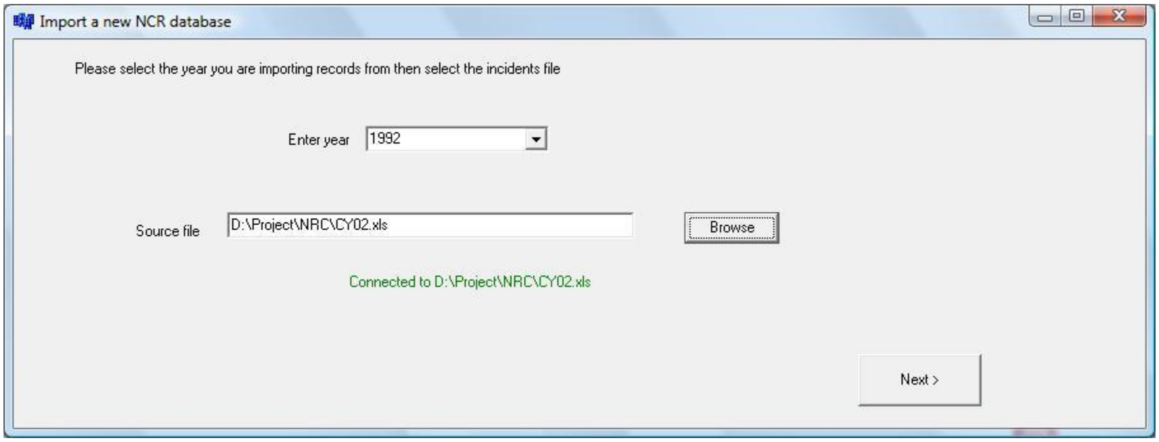


Figure 16: Input file selection window

A new window shown in **Figure 17** prompts the user to click "Import" in order to check the validity of the input file and extract the relevant data; A preview panel is displayed so the user can see a brief description of the current incident. By clicking "Migrate" the user triggers the actual migration which is done sequentially records by record and might take several minutes depending on the amount of data to be processed.

Figure 17 shows an ongoing migration process.

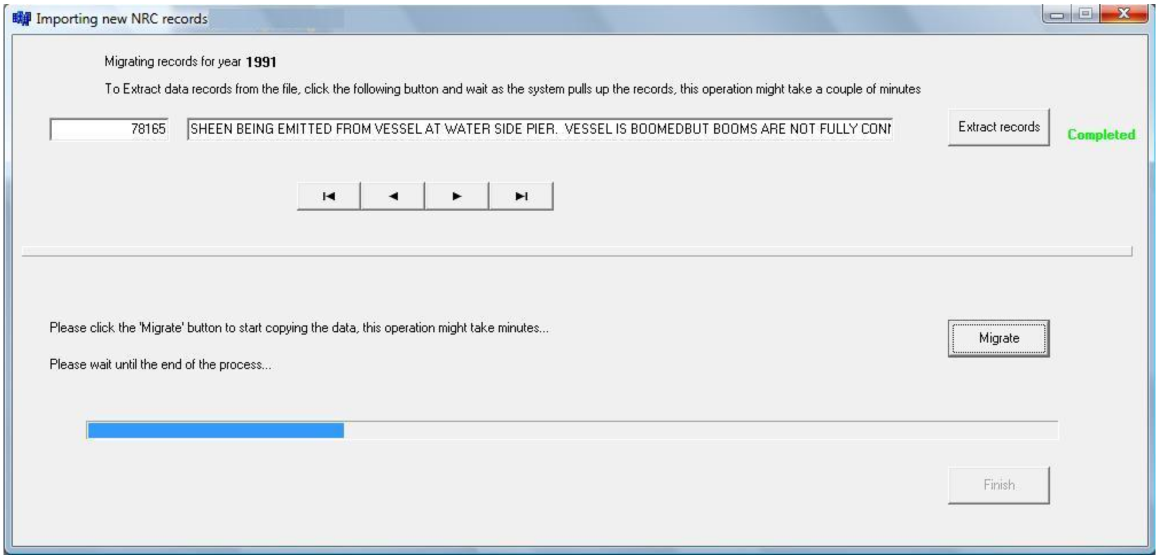


Figure 17: Program is importing records from the NRC file

A similar window is provided for the user who wants to import new incidents records from the HSEES database as shown in **Figure 18.**

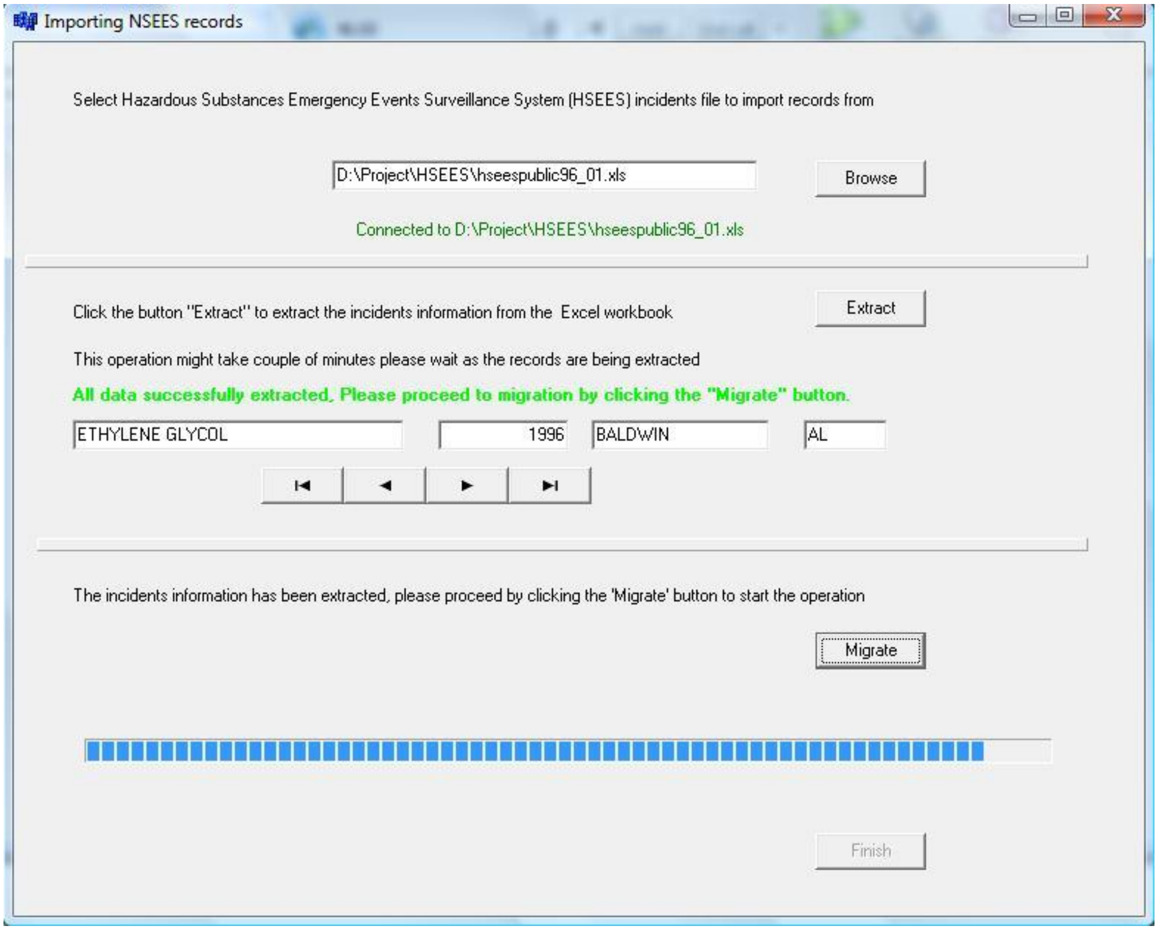


Figure 18: Program is importing records from the HSEES file

The following screen integrates the capacity of handling multiple types of queries based on the most common fields:

- Type of incident
- Incident year
- The state where it happened
- The county
- The chemical involved
- The number of people injured
- The number of people hospitalized
- The number of fatalities
- The damage amount

Based on this list of attributes, the user can create any possible query through the interface shown in **Figure 19**.

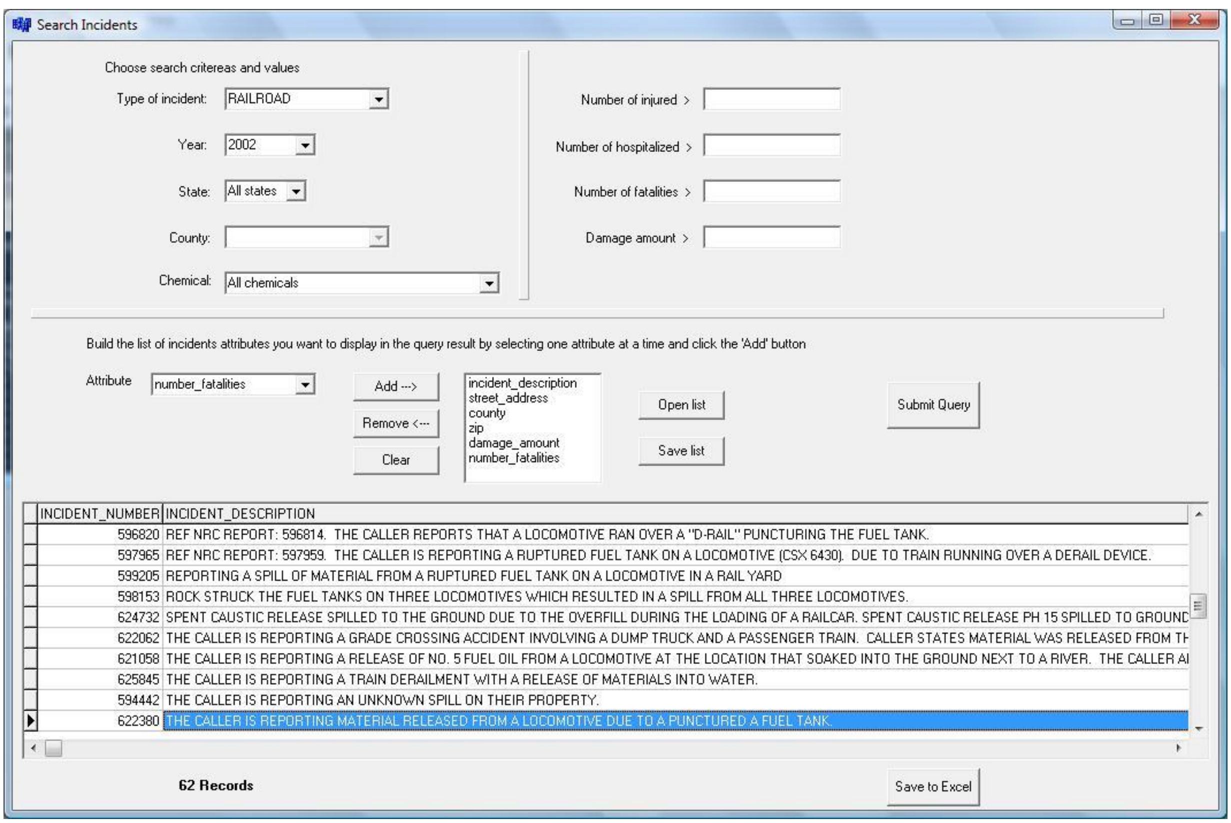


Figure 19: Query interface

Before submitting the query to the system, the user can construct a list of attributes among the 114 to be shown in the order he desires.

Along with a set of predefined attributes provided by default, the user can save his own list of attributes of his interest for future use.

When the "Submit Query" button is clicked, the result dataset displays in the form of table on the screen. The user can arrange the records in an ascending or descending order of the attribute of his choice. For example: the incidents that happened in Mississippi in 2005 and involved more than 5 injured ordered by damage amount in descending order. The user just clicks on the title of a column to sort the elements according to that particular attribute.

The user can export the results of his query to an Excel datasheet for printing or later analysis. The following screenshot in **Figure 20** shows a result of query that has been saved to an Excel file.

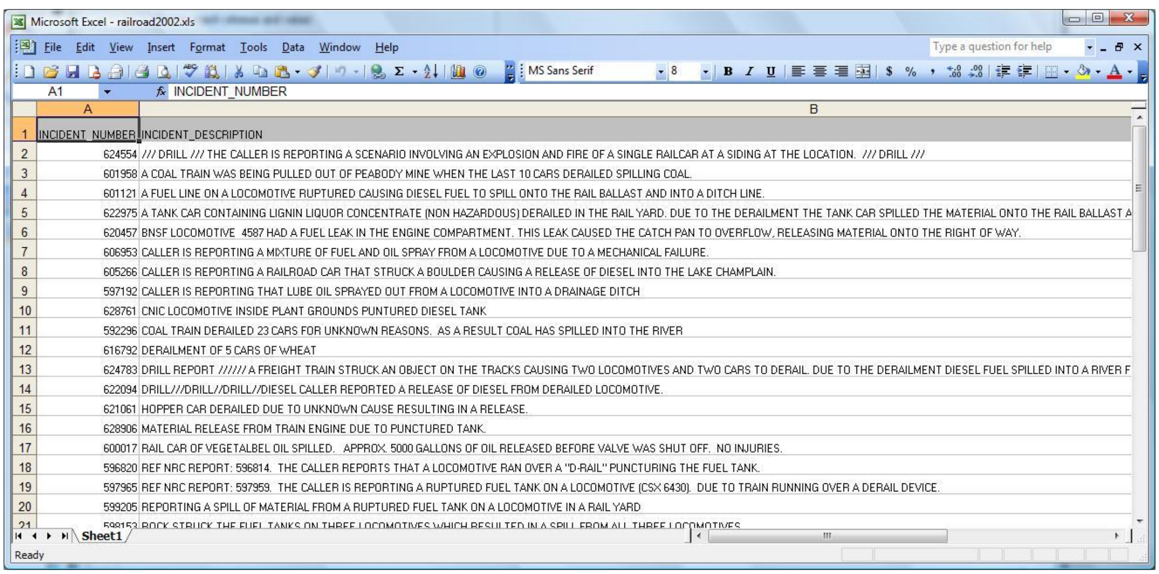


Figure 20: Exported query results

If the user is interested in one particular incident from the query result, he can click on "Lookup incident" from the main drop down menu and enter the incident number as shown in the following screenshot in **Figure 21**.

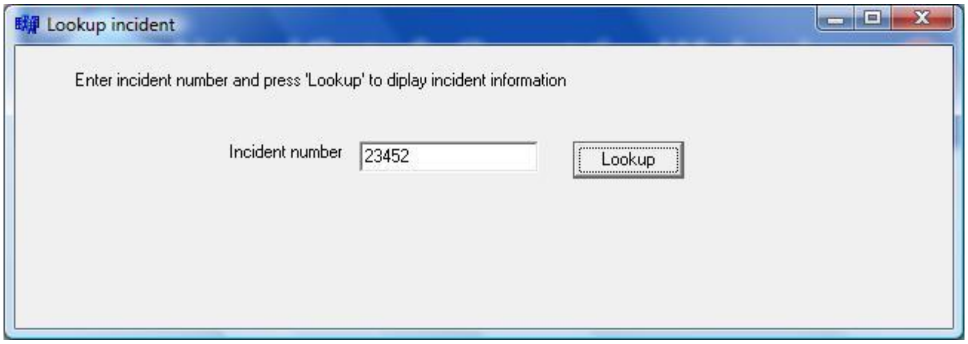


Figure 21: Search incident based on number

A printable incident report page will show on the screen with all the incidents attributes, **Figure 22** shows a report page print preview for incident number 222.

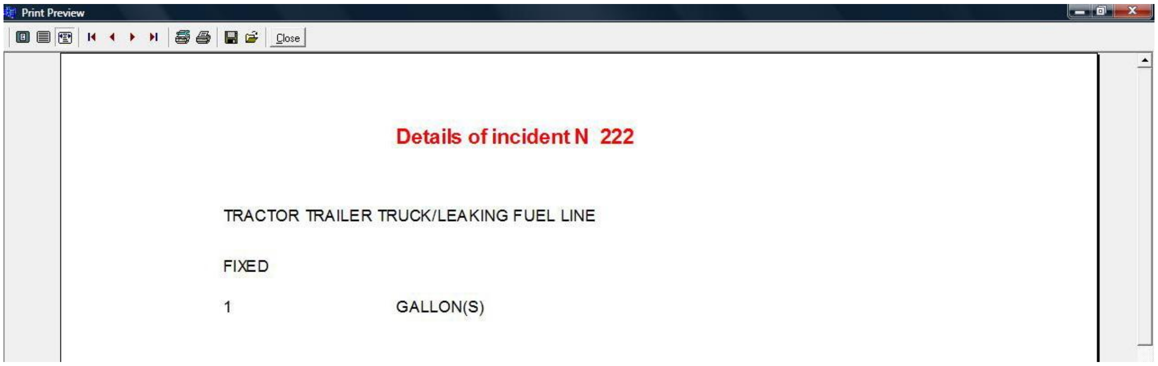


Figure 22: Incident details print preview

Conclusions

A chemical spill incident knowledge base has been developed, which provides quick access to the data. This system is flexible, portable and extensible. The knowledge-base framework allows users to:

- Import annual records from each of the agencies that hold the most detailed incidents information
- Adapt the system to any schema changes occurring in the input files
- Provide a structured format of data that is easy to manage and use
- Provide a very robust and secure DBMS providing a very increased performance
- A graphical user interface where users can construct any complex query without needing to master any query language
- Be able to save the environment setting in a profile file, so users do not need to go through the process of reconstructing their most used queries
- Perform sorting options on the resulting dataset to evaluate certain patterns in the data
- Export a result dataset in a rich format (Microsoft Excel file) for future analysis of the data
- Show and print an incident report that reassembles all the incident details into a two page, easy to read report.

With all these functionalities provided in a standalone version deliverable on one single CD-ROM, the knowledge-base meets all the requirements of this project.

Future Work

This project tackles a real world challenge and a big research theme. Several major companies are currently investing in data and services integration and as a result, many businesses are now migrating towards the loosely coupled architectures such as the

Service Oriented Architecture (SOA) (communication machine to machine is based on messages) [9].

Since the knowledge-base system has a consistent backend database schema, it would be more exploitable in the future if the database can be shared over a network and accessed via Web services from multiple locations. The client and server therefore will be using XML messages to communicate. This futuristic design will not only allow a larger number of concurrent queries to be supported but also allows clients to not have to depend on one rigid schema but can retrieve data in XML format and tailor it to their desired format.

Acknowledgements

This research was funded by US Department of Homeland Security (No. 4000055423). The suggestions and advices provided by Mr. Liang Huang, a graduate student of the University of Mississippi are highly appreciated

References

[1] The National Center for Hydro-science and Engineering. The University of Mississippi, *http://www.ncche.olemiss.edu, Jan 15, 2007*

[2] The National Response Center, *http://www.nrc.uscg.mil, Feb 2, 2007*

[3] Department of Health and Human Services, Agency for Toxic Substances and Disease Registry, *http://www.atsdr.cdc.gov, Apr 12, 2007*

[4] Simon Fraser University, Canada, *http://mis.bus.sfu.ca/tutorials/MSAccess/tutorials/intro.pdf, Mar 7, 2007*

[5] Oracle Corporation, *www.oracle.com, Apr 9, 2007*

[6] Sybase Enterprise, *http://www.sybase.com, Feb 15, 2007*

[7] FunctionX tutorials, *http://www.functionx.com/, Jun 20, 2007*

[8] Wikipedia, *http://www.wikipedia.com, Jul 5, 2007*

[9] International Business Machines, *www.ibm.com , Jan 9, 2007*

[10] Ian Abramson, Michael S. Abbey, and Michael Corey "*Oracle Database 10***g***: A Beginner's Guide*", McGraw-Hill 2004, ISBN 0-07-223078-9

[11] Ray Lischner, *"C++ in a Nutshell",* O'Reilly 2003, ISBN: 0-596-00298-X

[12] Lynne Dunkley and Andrew Smith, Improving Access of Commercial Software Developer to Formal Methods: Intergrating MERISE with Z, Proceedings of the Methods Integration Workshop, march 1996, UK

Appendix

Appendix 1: **Description of the data fields within the incidents chemical spill knowledge-base.**

No	Field Name	Field Description
1	Incident Number	Unique ID for each incident
2	Incident Description	Detailed explanation o the incident
3	Type of Incident	Specific type of incident being reported
4	Incident Date Time	Date and Time incident occurred
5	Incident DTG	Date Time Group - Discovered, Occurred or Planned
6	Incident Year	Year when the incident occurred
7	Incident Season	Season when the incident occurred
8	Incident Weekday	Portion of week when the incident occurred
9	Incident Time	Time range that the incident occurred
10	Incident Location	Descriptive explanation for the location of the incident
11	Location Street	Street address of the incident location
12	Location Nearest City	City or Town nearest to the incident location
13	Location State	State where incident occurred
14	Location County	County where incident occurred
15	Location Zip	Postal Zip code where incident occurred
16	Lat Deg	Degrees of Latitude for incident location
17	Lat Min	Minutes of Latitude for incident location
18	Lat Sec	Seconds of Latitude for incident location
19	Lat Quad	Latitude Quadrant for incident location
20	Long Deg	Degrees of Longitude for incident location
21	Long Min	Minutes of Longitude for incident location
22	Long Sec	Seconds of Longitude for incident location
23	Long Quad	Longitude Quadrant for incident location
24	Fire Involved	Indicates if any fire was involved with the incident
25	Any Evacuations	Indicates if any evacuations resulted from the incident
26	Number Evacuated	Approximate number of people evacuated
27	Radius of Evacuation	Indicates the approximate evacuation radius from the incident
28	Any Injuries	Indicates if any injuries occurred as a result of the incident
29	Number Injured	Number of people injured
30	Age_Rng1	number of people injured between birth and 19 years of age
31	Age_Rng2	number of people injured between 20 and 64 years of age
32	Age_Rng3	number of people injured 65 years of age
33	Inj_Tra	Number of people with trauma injuries

34	Inj_Resp	Number of people with respiratory system injuries
35	Inj_Eye	Number of people with eye irritation
36	Inj_Gastro	Number of people with gastrointestinal problems
37	Inj_Heat	Number of people with heat stress injuries
38	Inj_Chem	Number of people with chemical burn injuries
39	Inj_Therm	Number of people with thermal burn injuries
40	Inj_Skin	Number of people with skin irritation injuries
41	Inj_Cns	Number of people with dizziness or other CNS symptoms
42	Inj_Hache	Number of people with headaches
43	Inj_Hrt	Number of people with heart problems
44	Inj_Sob	Number of people with shortness of breath
45	Number Hospitalized	Number injured which required removal to an emergency facility
46	Sev_Hospr	Number injured where injury severity required treatment at hospital without being admitted or people was transported to hospitals for observation with no treatment
47	Sev_Nhosp	Number injured where injury severity required treatment on the scene (first-aid); or injured was seen by a private physician within 24 hrs; or injuries were experienced within 24 hrs of the incident and reported by an official
48	Any Fatalities	Indicates if any fatalities occurred as a result of the incident
49	Number Fatalities	Number of fatalities that occurred
50	Any Damages	Indicates if there was any property damage as a result of the incident
51	Damage Amount	Monetary estimation of total property damage
52	**Medium Description**	Medium affected as a result of the incident
53	Additional Medium Info	More detailed description of medium affected
54	Residential Area	If there is residential area within 1/4 mile of incident
55	Body of Water	Immediate Body of water impacted by the incident if applicable
56	Tributary Of	Additional Body of water the impact would flow into if applicable
57	Nearest River Mile Marker	The nearest river mile marker of the incident location if applicable
58	Release Secured	Indicates if the release of material has been secured
59	Estimated Duration of Release	Length of time the release of material occurred
60	Duration Unit	Unit of measure for the estimated duration of release
61	Release Rate	Rate of the material release
62	Release Rate Unit	Unit of measure for the release rate

63	Desc Remedial Action	Description of what was being done to mitigate the material release
64	Water Supply Contaminated	Indicates if a drinking water source was contaminated by the release
65	Sheen Size	If a sheen on the water, approximate size of sheen
66	Sheen Size Units	Unit of measure used to indicate the sheen size
67	Sheen Size Length	Approximate length of the sheen
68	Sheen Size Length Units	Unit of measure for the Sheen Size Length
69	Sheen Size Width	Approximate width of the sheen
70	Sheen Size Width Units	Unit of measure for the Sheen Size Width
71	Offshore	Indicates if the incident location is offshore
72	Sheen Color	General color of sheen
73	Sheen Odor Description	Description of any odor related to the sheen
74	Water Temperature	Approximate temperature of the water affected by the release
75	Additional Info	Any additional information not covered elsewhere in the report
76	State Agency on Scene	Name of any state agency which was on the incident scene
77	State Agency Notified	Any State agency already notified
78	Federal Agency Notified	Any Federal agency notified besides the NRC
79	NPDES	National Pollutant Discharge Elimination System identifier
80	NPDES Compliance	Indicates if discharge was in compliance with NPDES
81	Exposed Underwater	Indicates if the underwater pipeline was uncovered or exposed
82	Continuous Release Type	Type of Continuous Release being reported
83	Chemical 1	Name of No.1 chemical released
84	Amount Of Chemical 1	Amount of Chemical 1 Released
85	Chm_Qcat1	Range for the amount of Chemical 1
86	Unit Of Chemical 1	Unit of Measure for Amount Released of Chemical 1
87	**If Reached Water**	Indicates if Chemical 1 reached a body of water
88	Rels1chem1	Type of release for Chemical 1
89	Amount In Water	Amount of Chemical 1 that reached water
90	Unit of Measure Reach Water	Unit of Measure for Amount in Water of Chemical 1
91	Upper Bounds	Maximum Amount allowed under Continuous Release Permit of Chemical 1
92	Upper Bounds Unit	Unit of measure used for Upper Bounds of Chemical 1
93	Upper Bounds Rate	Maximum rate of release allowed under Continuous Release Permit of Chemical 1
94	Chemical 2	Name of No.2 chemical released
95	Chm_Qcat2	Range for the amount of Chemical 2

96	Unit Of Chemical 2	Unit of Measure for Amount Released of Chemical 2
97	Rels1chem2	Type of release for Chemical 2
98	Chemical 3	Name of No.3 chemical released
99	Chm_Qcat3	Range for the amount of Chemical 3
100	Unit Of Chemical 3	Unit of Measure for Amount Released of Chemical 3
101	Rels1chem3	Type of release for Chemical 3
102	Chemical 4	Name of No.4 chemical released
103	Chm_Qcat4	Range for the amount of Chemical 4
104	Unit Of Chemical 4	Unit of Measure for Amount Released of Chemical 4
105	Rels1chem4	Type of release for Chemical 4
106	Chemical 5	Name of No.5 chemical released
107	Chm_Qcat5	Range for the amount of Chemical 5
108	Unit Of Chemical 5	Unit of Measure for Amount Released of Chemical 5
109	Rels1chem5	Type of release for Chemical 5
110	Chemical 6	Name of No.6 chemical released
111	Chm_Qcat6	Range for the amount of Chemical 6
112	Unit Of Chemical 6	Unit of Measure for Amount Released of Chemical 6
113	Rels1chem6	Type of release for Chemical 6
114	Counter Measures	Measures taken to mitigate the incident

Appendix 2: SQL data definition code for creating the schema in Oracle 10 g

```
/*==============================================================*/
/* Table: CHEMICAL                                              */
/*==============================================================*/
create table CHEMICAL  (
   CHEMICAL_ID          INTEGER                         not null,
   CHEM_NAME            VARCHAR(70),
   constraint PK_CHEMICAL primary key (CHEMICAL_ID)
);

/*==============================================================*/
/* Table: INCIDENTS                                             */
/*==============================================================*/
create table INCIDENTS  (
   INCIDENT_NUMBER      NUMBER(7)                       not null,
   INCIDENT_DESCRIPTION VARCHAR(240),
   TYPE_OF_INCIDENT     VARCHAR(40),
   INCIDENT_DATE_TIME   VARCHAR(30),
   INCIDENT_DTG         VARCHAR(10),
   INCIDENT_YEAR        NUMBER(4),
   INCIDENT_SEASON      VARCHAR(30),
   INCIDENT_WEEKDAY     VARCHAR(10),
   INCIDENT_TIME        VARCHAR(8),
   INCIDENT_LOCATION    VARCHAR(80),
   STREET_ADDRESS       VARCHAR(120),
   COUNTY               VARCHAR(30),
   NEAREST_CITY         VARCHAR(40),
   STATE                VARCHAR(2),
   ZIP                  NUMBER(5),
   LAT_DEG              NUMBER(2),
   LAT_MIN              NUMBER(2),
   LAT_SEC              NUMBER(2),
   LAT_QUAD             NUMBER(2),
   LONG_DEG             NUMBER(2),
   LONG_MIN             NUMBER(2),
   LONG_SEC             NUMBER(2),
   LONG_QUAD            NUMBER(2),
   FIRE_INVOLVED        VARCHAR(1),
   NUMBER_EVACUATED     NUMBER(5),
   RADIUS_OF_EVACUATION NUMBER(7,2),
   NUMBER_INJURED       NUMBER(5),
   AGE_RNG1             NUMBER(5),
   AGE_RNG2             NUMBER(5),
   AGE_RNG3             NUMBER(5),
   NUMBER_HOSPITALIZED  NUMBER(5),
   SEV_HOSPR            NUMBER(5),
   SEV_NHOSP            NUMBER(5),
   NUMBER_FATALITIES    NUMBER(5),
   DAMAGE_AMOUNT        NUMBER(9),
   MEDIUM_DESCRIPTION   VARCHAR(22),
   ADDITIONAL_MEDIUM_INFO VARCHAR(80),
   RESIDENTIAL_AREA     VARCHAR(1),
   BODY_OF_WATER        VARCHAR(40),
   TRIBUTARY_OF         VARCHAR(40),
   NEAREST_RIVER_MILE_MARKER NUMBER(6,2),
   RELEASE_SECURED      VARCHAR(1),
   ESTIMATED_DURATION_OF_RELEASE VARCHAR(12),
   DURATION_UNIT        VARCHAR(40),
   RELEASE_RATE         VARCHAR(40),
   RELEASE_RATE_UNIT    VARCHAR(40),
```

```
  DESC_REMEDIAL_ACTION VARCHAR(240),
  WATER_SUPPLY_CONTAMINATED VARCHAR(1),
  SHEEN_SIZE          VARCHAR(40),
  SHEEN_SIZE_UNITS     VARCHAR(40),
  SHEEN_SIZE_LENGTH    NUMBER(5),
  SHEEN_SIZE_LENGTH_UNITS VARCHAR(40),
  SHEEN_SIZE_WIDTH     VARCHAR(5),
  SHEEN_SIZE_WIDTH_UNITS VARCHAR(40),
  OFFSHORE            VARCHAR(1),
  SHEEN_COLOR         VARCHAR(40),
  SHEEN_ODOR_DESCRIPTION VARCHAR(40),
  WATER_TEMPERATURE    NUMBER(2),
  ADDITIONAL_INFO      VARCHAR(500),
  STATE_AGENCY_ON_SCENE VARCHAR(40),
  STATE_AGENCY_NOTIFIED VARCHAR(40),
  FEDERAL_AGENCY_NOTIFIED VARCHAR(40),
  NPDES               VARCHAR(10),
  NPDES_COMPLIANCE     VARCHAR(1),
  EXPOSED_UNDERWATER   VARCHAR(1),
  CONTINUOUS_RELEASE_TYPE VARCHAR(40),
  AMOUNT_IN_WATER      NUMBER(11,2),
  UNIT_OF_MEASURE_REACH_WATER VARCHAR(40),
  UPPER_BOUNDS        VARCHAR(20),
  UPPER_BOUNDS_UNIT    VARCHAR(40),
  UPPER_BOUNDS_RATE    VARCHAR(40),
  REPORTS_LINK        VARCHAR(240),
  SEVERITY_LEVEL       NUMBER(2),
  constraint PK_INCIDENTS primary key (INCIDENT_NUMBER)
);

/*==============================================================*/
/* Table: INJURED                                               */
/*==============================================================*/
create table INJURED  (
  INCIDENT_NUMBER      NUMBER(7)                    not null,
  INJURIY_ID          NUMBER(3)                   not null,
  NUMBER_OF_PEOPLE     INTEGER,
  constraint PK_INJURED primary key (INCIDENT_NUMBER, INJURIY_ID)
);

/*==============================================================*/
/* Index: INJURED2_FK                                           */
/*==============================================================*/
create index INJURED2_FK on INJURED (
  INJURIY_ID ASC
);

/*==============================================================*/
/* Table: INJURIES                                              */
/*==============================================================*/
create table INJURIES  (
  INJURIY_ID          NUMBER(3)                   not null,
  INJURY_TYPE         CHAR(30),
  constraint PK_INJURIES primary key (INJURIY_ID)
);

/*==============================================================*/
/* Table: RELEASED                                              */
/*==============================================================*/
create table RELEASED  (
  CHEMICAL_ID         INTEGER                     not null,
  INCIDENT_NUMBER      NUMBER(7)                    not null,
```

```
  AMOUNT              INTEGER,
  UNIT_AMOUNT         VARCHAR(20),
  TYPE_OF_RELEASE     VARCHAR(20),
  constraint PK_RELEASED primary key (CHEMICAL_ID, INCIDENT_NUMBER)
);

/*==============================================================*/
/* Index: RELEASED2_FK                                          */
/*==============================================================*/
create index RELEASED2_FK on RELEASED (
  INCIDENT_NUMBER ASC
);

alter table INJURED
  add constraint FK_INJURED_INJURED_INCIDENT foreign key (INCIDENT_NUMBER)
    references INCIDENTS (INCIDENT_NUMBER);

alter table INJURED
  add constraint FK_INJURED_INJURED2_INJURIES foreign key (INJURIY_ID)
    references INJURIES (INJURIY_ID);

alter table RELEASED
  add constraint FK_RELEASED_RELEASED_CHEMICAL foreign key (CHEMICAL_ID)
    references CHEMICAL (CHEMICAL_ID);

alter table RELEASED
  add constraint FK_RELEASED_RELEASED2_INCIDENT foreign key (INCIDENT_NUMBER)
    references INCIDENTS (INCIDENT_NUMBER);
```

Attachment A