Volume 12, No. 5
August 2014
ISSN: 1545-679X

INFORMATION SYSTEMS EDUCATION JOURNAL

Special Issue: Teaching Cases

In this issue:

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **EDSIG**, the Education Special Interest Group of AITP, the Association of Information Technology Professionals (Chicago, Illinois). Publishing frequency is six times per year. The first year of publication is 2003.

ISEDJ is published online (http://isedjorg) in connection with ISECON, the Information Systems Education Conference, which is also double-blind peer reviewed. Our sister publication, the Proceedings of ISECON (http://isecon.org) features all papers, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is about 45%.

Information Systems Education Journal is pleased to be listed in the 1st Edition of Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org.

2014 AITP Education Special Interest Group (EDSIG) Board of Directors

Wendy Ceccucci
Quinnipiac University
President – 2013-2014

Scott Hunsinger
Appalachian State Univ
Vice President

Alan Peslak
Penn State University
President 2011-2012

Jeffry Babb
West Texas A&M
Membership Director

Michael Smith
Georgia Institute of Technology
Secretary

George Nezlek
Univ of North Carolina
Wilmington -Treasurer

Eric Bremier
Siena College
Director

Nita Brooks
Middle Tennessee State Univ
Director

Muhammed Miah
Southern Univ New Orleans
Director

Leslie J. Waguespack Jr
Bentley University
Director

Peter Wu
Robert Morris University
Director

S. E. Kruck
James Madison University
JISE Editor

Nita Adams
State of Illinois (retired)
FITE Liaison

Copyright © 2014 by the Education Special Interest Group (EDSIG) of the Association of Information Technology Professionals (AITP). Permission to make digital or hard copies of all or part of this journal for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial use. All copies must bear this notice and full citation. Permission from the Editor is required to post to servers, redistribute to lists, or utilize in a for-profit or commercial use. Permission requests should be sent to Nita Brooks, Editor, editor@isedj.org.

Information Systems Education Journal

Editors

Nita Brooks
Senior Editor
Middle Tennessee State University

Thomas Janicki
Publisher
University of North Carolina Wilmington

Donald Colton
Emeritus Editor
Brigham Young University Hawaii

Jeffry Babb
Associate Editor
West Texas A&M University

Wendy Ceccucci
Associate Editor
Quinnipiac University

Melinda Korzaan
Associate Editor
Middle Tennessee State University

George Nezlek
Associate Editor
Univ of North Carolina Wilmington

Samuel Sambasivam
Associate Editor
Azusa Pacific University

Anthony Serapiglia
Teaching Cases Co-Editor
St. Vincent College

Lawrence Cameron
Teaching Cases Co-Editor
University of Montana

ISEDJ Editorial Board

Samuel Abraham
Siena Heights University

Teko Jan Bekkering
Northeastern State University

Gerald DeHondt II

Janet Helwig
Dominican University

Scott Hunsinger
Appalachian State University

Mark Jones
Lock Haven University

James Lawler
Pace University

Michelle Louch
Duquesne University

Cynthia Martincic
Saint Vincent College

Muhammed Miah
Southern Univ at New Orleans

Marianne Murphy
North Carolina Central University

Alan Peslak
Penn State University

Bruce Saulnier
Quinnipiac University

Li-Jen Shannon
Sam Houston State University

Karthikeyan Umapathy
University of North Florida

Bruce White
Quinnipiac University

Peter Y. Wu
Robert Morris University.

Teaching Case

Cleaning Data Helps Clean the Air

Kelley Donalds
kelley.donalds@bridgew.edu

Xiangrong Liu
Xiangrong.liu@bridgew.edu

Bridgewater State University
Bridgewater, MA

Abstract

In this project, students use a real-world, complex database and experience firsthand the consequences of inadequate data modeling. The U.S. Environmental Protection Agency created the database as part of a multimillion dollar data collection effort undertaken in order to set limits on air pollutants from electric power plants. First, students explore the database to identify design limitations from the perspective of a data analyst with a specific goal. Second, students create a new database design which overcomes identified problems. Through this case study, students develop the skill to infer usage implications by studying the design of an existing database. This is important since developers often inherit databases designed by others. Students also learn how to prepare data stored in a relational database for a data analysis project. By experiencing the consequences of an inadequate design from a user perspective, students can better appreciate the importance of relational database design principles and become more committed to using them.

Keywords: database design, data modeling, data cleaning, referential integrity, normalization

1. INTRODUCTION

John and Kayla had just started their new jobs as Data Analysts at the Utility Research Institute (URI), a non-profit organization that conducts research on behalf of its funding organizations -- primarily electric utility companies operating within the United States. John had an M.S. in Computer Science and had worked as a Database Management Administrator for the past five years. Kayla had just graduated with an M.S. in Mathematics with a concentration in Statistics. They had been assigned to work together on a project analyzing data that was compiled by the U.S. Environmental Protection Agency (EPA). The data was collected as part of a process for establishing the first ever national standards limiting emissions of hazardous air pollutants such as mercury from coal and oil fired power plants. The EPA had made the data available to the public in the form of a Microsoft Access database and the Institute wanted to use this data to determine boiler features and pollution control equipment that would satisfy emission standards for all of the newly regulated pollutants.

2. BACKGROUND

To kick off the project, Kayla and John's manager, Ravi, briefed them on the regulatory history of air pollutants within the U.S. utility industry. He said that the recent 2011 ruling, known as the Mercury and Air Toxics Rule (MATR), imposed the first ever national limits on heavy metals such as mercury and acid gas emissions from coal and oil power plants (EPA 2011b). The rule specifically limited air emissions of mercury, filterable particulate matter and hydrochloric acid from coal and oil fired plants with at least 25 megawatt hours of generating capacity. The EPA had decided to use filterable particulate matter and hydrochloric acid as surrogates for all non-mercury metals and all acid gases respectively. To show the importance of the project, Ravi shared an article from a trade journal (Neville 2012) in which industry representatives described MATR as the most expensive regulation under the Clean Air Act (CAA) in terms of direct costs and the most extensive intervention into the power market that the EPA had ever attempted. EPA's own detailed analysis estimated that the rule would affect about 500 coal-fired plants and 100 oil fired plants at an annual cost of $9.6 billion (EPA 2011a). Given the significant compliance costs, it was likely that some utilities would be making "invest or retire" decisions for many plants -- especially older ones.

John asked Ravi why the electric utilities hadn't been subject to earlier regulation of these air pollutants. Ravi explained that while electric utilities were no stranger to regulation under the CAA, they had been treated differently than other industries in the major 1990 amendments (EPA 2013a). Congress passed these major revisions to better control urban air pollution (Title I), pollutants from mobile sources (Title II), toxic air emissions (Title III), acid rain (Title IV) and ozone-depleting chemicals (Title VI). Title V delegated responsibility for regulatory oversight to individual states via a permitting process. Title IV had imposed significant regulations on the utility industry to better control emissions of sulfur dioxide which contributes to acid rain. Title I had imposed limits on emissions of nitrous oxides and particulate matter which contribute to urban area smog and also impacted the utility industry.

However, Ravi explained that the electric utility industry had successfully forestalled regulation under Title III of the amendments (e.g. Lemonick 1990). Title III listed 189 air toxins for which the EPA was required to identify source categories that would be subject to future regulation under section 112 of the CAA. Standards under section 112 were based on what was referred to as maximum achievable control technology (MACT). For existing sources, MACT sets a minimum level of stringency called the floor which is the average emission "achieved by the best performing twelve percent of existing sources in the category or the best performing five sources for source categories with less than thirty sources" (EPA 2013b). Quoting the CAA, congress had required the EPA to perform a study of the "hazards to public health reasonably anticipated to occur" as a result of emissions of listed air toxins and to regulate electric utilities under Title III *only* "if the Administrator finds such regulation is appropriate and necessary after considering the results of the study" (EPA 2013c). A general report regarding all the listed air toxins by utilities was due in three years and an additional report addressing health effects of mercury emissions from utilities and other industries was due in four years.

Kayla asked why congress had given utility companies a reprieve; it didn't seem to make sense if they were significant sources of the listed air toxins. Ravi surmised that congress may have been more lenient with utility companies under Title III since they were already primary targets of regulation under Titles I and IV of the 1990 amendments. Both John and Kayla were surprised that emissions of heavy metals such as mercury, arsenic and lead had never been regulated within the utility industry. Noting that some individual states did limit power plant emissions of heavy metals such as mercury, Ravi agreed that it was surprising that so many air toxins from power plants had not been regulated at the federal level – at least until now.

Mercury, in particular, had received significant attention (e.g. EPA 1997, Center for Disease Control 1999, Physicians for Social Responsibility 2004). As explained in the 1997 EPA report, mercury released by industrial sources into the air can circulate in the atmosphere for up to a year and can be deposited on land and water thousands of miles from the original source. When heavy metal mercury is consumed by living organisms, it is converted to bioaccumulative methyl-mercury which becomes more concentrated in organisms higher in the food chain. A fact sheet issued by the Physicians for Social Responsibility (2004) describes mercury as a "potent neurotoxin" that affects the functioning of the central nervous system and explains that most Americans are exposed to mercury through the consumption of fish – especially of higher food chain predatory fish like swordfish and tuna. In its 1997 report, the EPA had estimated annual emissions of mercury within the U.S. to be about 158 tons of which 87% came from waste and fossil fuel combustion. However, since waste combustion had been subject to earlier regulation, fossil fuel combustion (primarily coal) was now the dominant source of mercury emissions in the United States. Ravi summed up the discussion by stating that two decades after the 1990 amendments, the 2011 MATR had listed electric utilities as a source category under section 112 of the CAA and that the long delay was a result of years of litigation between industry, non-governmental organizations, states and the EPA.

3. RESEARCH PURPOSE

The discussion then switched to the purpose of the research and the EPA data. Ravi explained that in order to gather the data needed to set the standards, the EPA had issued a two-phase information collection request in 2009 (EPA 2009). In the first phase, electric generating units (EGUs) subject to the new regulation completed a twenty-five page paper survey providing the most recent twelve months of emissions test and fuel analysis data since 2005 as well as data about plant equipment (e.g. boiler characteristics, pollution controls) and permitting requirements. In the second phase, the EPA selected EGUs who were believed to be the best performing units within specified pollutant categories. These EGUs were required to conduct emissions stack testing to measure flue gas entering the atmosphere and to conduct analyses of fuel used during testing. The cost of data collection and quality assurance was estimated to be about $10 million and the cost of stack and fuel testing was estimated to be about $66 million (EPA 2009). In order to leverage this investment, the Institute wanted to gain as much knowledge as possible from the EPA data which was made available to the public in the form of two MS Access databases -- one for each collection phase. They would start with the data from the first collection phase. Ravi was sure that this task alone would be very challenging. After they had mastered the Phase I database, they would consider integration of Phase II data. Links to the original data and descriptive information are provided in Table 1.

The purpose of the current project is to determine which combinations of equipment provide the best overall control of multiple pollutants. Certain boilers can remove pollutants during combustion or while coal is being burned. For example, fluidized bed boilers float and tumble burning coal on upward jets of air. The tumbling allows solids such as limestone to be mixed in with the coal and absorb pollutants such as sulfur dioxide. Other boilers are designed to burn coal at lower temperatures which inhibit the formation of nitrous oxides. In addition, different types of post combustion controls can remove pollutants from the flue gas before it is released into the air through the smokestack.

Kayla had one nagging question: What was the value of analyzing equipment that wasn't intended to control emissions of the newly regulated pollutants? Ravi explained that the EPA (2011a) had argued that the new standards were based on "existing, commercially proven technologies that are...frequently used in this industry such as electrostatic precipitators, fabric filters (bag houses), flue gas desulfurization (scrubbers) or dry sorbent injection." In other words, equipment used to control sulfur dioxide, nitrous oxides, particulate matter also controlled emissions of the newly regulated pollutants – at least according to the EPA. Indeed as Ravi pointed out, the EPA was using particulate matter as a proxy for all non-mercury metals. "So does this mean, the newly regulated pollutants were – in effect -- already being regulated" Kalya asked? Ravi wasn't so sure stating that "these are the kinds of questions we need to answer with our research" and that "controls for different pollutants may interact in ways that do not simultaneously reduce all regulated pollutants".

4. A DATA NARRATIVE

John had spent the last week studying the Phase I EPA database and was meeting with Kayla to give her an overview of what he had learned so far. He also wanted to get a better understanding of what data and what format would be required to conduct statistical analyses. Referring to the EPA database diagram, John convinced Kayla that the EPA Phase I was complex involving many dimensions. It contained forty different tables which were linked together by almost as many relationships. He showed her a sketch (Figure 1) of the data entity relationships which he had created based on the EPA database diagram.

In order to get a better understanding of the content of the database, John explained to Kayla that he had created a smaller "test" M.S. Access database by deleting some of the tables and fields from the first phase EPA database. He believed that the smaller database contained the most important data for their research project and that the simplification would facilitate their preliminary analysis. All relationships were those created by the EPA and no records had been deleted from the remaining tables. A screen shot of MS Access relationships in the test database is shown in Figure 2.

John had many questions but would do his best to explain the Figure 2 diagram to Kayla. A facility, described in the `facility_information` table, is all the property, plant and equipment that resides at single geographic location and that has a legal owner. A configuration is a set of equipment components ordered by their physical location within the electricity generation process. A facility can have multiple configurations which are possibly operated concurrently at a given point in time or possibly which have changed over time due to the addition, modification or removal of particular equipment. Configurations are described in the `configuration_components` table.

Each configuration starts with one or more of what was labeled as a "unit". Each unit is in turn mapped to one or more boilers in the `unit_boilers` table and boilers are described in the `boiler_ information` table. John knew that the information in the `boiler_information` table would be important but he did not know what a "unit" represented. It seemed that the label "unit" was so generic that it could represent any kind of equipment. Question 15 of the EPA survey required "identification (or designation) of all coal- and oil-fired steam generating units (boilers) (as defined by Clean Air Act section 112(a)(8)) located at this facility" The question parenthetically indicates that a steam generating unit is a "boiler" and a footnote indicates that either a `Boiler ID` or a `Generator ID` can be provided:

> Boiler ID as reported on U.S. DOE/EIA Form EIA-860 (2007), "Annual Electric Generator Report", schedule 6, part A, line 1, page 53 OR on schedule 6, part B, line 1, page 54 OR Generator ID as reported on "U.S. DOE/EIA From EIA-923 (2008)"

John wondered whether allowing the interchangeable use of boiler and generator ids was a source of design problems. According to the language of the CAA, the EPA is required to regulate steam generating units and the CAA defines an "electric utility steam generating unit'' as "any fossil fuel fired combustion unit of more than 25 megawatts that serves a generator that produces electricity for sale". Based on this definition, the steam-generating unit is not the same as the generator that produces electricity. The former generates steam and the latter generates electricity. Like the term "unit", "generate" also had multiple meanings. Further adding to the confusion, the term "steam" was often use to describe a generator as indicated on EIA (Energy Information Administration) Form 860:

> Enter the identification (ID) code for each boiler that provides steam to each combustible-fuel steam generator … and for each combined cycle steam turbine generator. Boilers may be associated with multiple generators.

It was also apparent that there is a many-to many relationship between boilers and generators. In order to clarify the terminology, John conducted some research and settled on the following definitions:

> A boiler is a vessel which burns fuel to boil water and create expanding, pressurized steam which is transferred to at least one turbine. The thermal energy will be converted into rotating kinetic energy.
>
> A turbine is a rotor with blades that is connected to the shaft of a generator. It uses rotary motion to convert kinetic to mechanical energy.
>
> A generator is copper wire coiled around a shaft that is surrounded by a giant magnet. When the shaft is rotated, electric current is created on the wire, converting mechanical energy to electrical energy.

He was confident that these definitions provided much needed semantic clarity. And he had also discovered that the qualifier "steam" was used to distinguish the type of turbine which in addition to steam included water, wind and gas types.

Based on the survey instructions, `unit_id` was possibly meant to refer to a generator but he was still unsure. In the entire phase I database, there was NO additional information stored about units beyond the id itself. He was puzzled why the configuration table included units and not boilers. CAA regulatory rules apply to boilers and not generators. This issue required further research; he had a nagging concern that it would be a cause of problems for their research.

Each configuration also has at least one chimney – called a stack – where gas exits the process. One or more pollution control devices may be installed after the unit and before a stack. The database contained four major groups of such post-combustion controls devices including particulate matter (PM) controls, nitrous oxide (NOx) controls, sulfur dioxide (SO2) and other controls. The "other" category contained mercury (Hg) control devices and Kayla and John agreed these would need to be separated. Control devices which are relatively independent (e.g. can be removed and relocated within a configuration or installed within another configuration) are referred to as "facility" controls and are described in the `facility_controls` table. In addition, boilers have design features to control NOx pollution which are described in the `boiler_nox_control` table. Air is sampled through ducts called sampling ports which can be placed at different locations within the process as well as at the exhaust stack. John noted that only pollution controls that were located *upstream* of (e.g. before) the sampling location should be associated with pollutant measurements at that location.

In the survey, utilities provided historical emissions data in the form of test reports. Each test report often corresponded to a compliance reporting requirement and each report in turn consists of multiple sampling runs where measurement devices collect and analyze samples of air during a discrete period of time. Multiple sampling runs might be used to ensure that measurements reflect steady state conditions of the electricity generation process. Each sampling run is in turn associated with one or more pollutants for which emissions are reported. The database contained emissions data for 106 different pollutants -- although many of these were infrequently reported. Kayla and John decided to focus only on the following pollutants: filterable particulate matter, sulfur dioxide (SO2), nitrogen oxide (NOx), total mercury (Hgt) and hydrogen chloride (HC1). John was initially confused about which type of mercury he should use but he had verified that total mercury is the sum of elemental mercury (Hg0), particulate bound mercury (Hgp) and oxidized mercury (Hg++). So for now, they would extract only Hgt. To further complicate matters, emissions were reported using different units of measurements including emissions rates (e.g. weight emitted per time period), emission factors (weight per heating fuel content) and concentrations (parts per air volume). These units of measurement are interdependent in that one may be derived from others given additional data. Kayla had done some initial investigation on converting emissions to a common unit of measurement and found that it was not straightforward. There were multiple conversion formulas which each made different assumptions and required different additional data. So to begin their analysis, Kayla and John agreed to use only sampling runs which reported emissions factors as pounds per million British Thermal Unit (lb/MMBtu) since this was the most frequent unit of measurement in the `sampling_run_pollutants` table. Emissions data is contained in `test_reports`, `sampling_runs` and `sampling_run_pollutants` tables.

5. THE EXTRACT

John wanted to know what data format would be required for statistical analysis. Kayla explained that the typical input for statistical software is a two dimensional file or table. Each row represents an observation and each column represents a variable. John referred to this type of input as a flat "denormalized" table. Kayla continued explaining that it is usual in statistical analyses that some variables are dependent (those to be predicted or explained) and others are independent (those that form the basis for explanation or prediction). Computers scientists might more easily understand dependent and independent variables as output and input variables. In the current project, dependent variables are pollutant emissions and independent variables are boiler characteristics and control equipment.

They needed to determine the unit of analysis or observation and tentatively decided to define the observation as a unique combination of boiler characteristics and pollution controls at a particular facility. They would average pollutant emissions to this level of analysis. John suspected there might be some situations where emissions measurements could not be unambiguously associated with unique equipment and in these cases the emissions data should be excluded from the analysis. It was also important that a single emissions measurement was not averaged into multiple observations since this would bias results by weighting some measurements more heavily than others. Also, multiple configurations of identical equipment at a specific facility should be merged into a single observation.

The discussion switched to data types – which was more straightforward than level of aggregation. Kayla suggested coding `boiler_firing_type` as a categorical data type with the following possible values: tangential, wall, cyclone, fluidized bed, integrated gas combustion cycle (IGCC) and other. Although not all statistical procedures handled categorical data types, initially she would conduct descriptive analyses by `boiler_firing_type`. The mapping of specific boiler firing type values which exist in the EPA database to the extract categories is shown equipment classification hierarchy shown in Table 3. For example, "front wall", "rear wall", "opposed wall", and "other" boiler firing types should be mapped to "wall" firing type.

Since a configuration can have a varying number of controls within a single category, Kayla suggested coding facility and boiler pollution controls as Boolean data types with 1 indicating presence of the control and 0 otherwise. Kayla and John drafted a preliminary structure for a data extract shown in Table 2. Like boiler firing types, the equipment classification hierarchy in Table 3 maps specific controls to the general controls in the extract file.

6. PRELIMINARY DATA ANALYSIS

John had some concerns about possible data anomalies which would affect the integrity of the data used for their research project. Their goal was to unambiguously relate emissions measurements to boiler characteristics and control equipment that was operational at the time of the test. He recognized that parts of the database did not meet normalization principles and some referential integrity constraints were missing. He came up with a plan to systematically investigate these issues. First, he would manually try to create extract records for some sample facilities. He had successfully done this for facility 663. MS-Access screenshots and the extract records for this facility are shown in figures 3 and 4 respectively. He had identified five additional facilities which he thought might present problems and would manually try to construct extract data for these facilities. The identifiers for the test facilities are: 56, 898, 1073, 1507 and 2324. For example, a potential problem for facility 1073 is that units 1 and 2 were each mapped to four boilers (1-4) in the `unit_boilers` table. He was concerned that the boilers would have different characteristics and had begun researching this plant using data at the Energy Information Administration web site. He had learned that in fact only boilers 1 and 2 should both be mapped to units 1 and 2. Second, in the process of creating extract data for the five facilities, he would make a list of problems in terms of relating emissions to equipment data. Third, he would design a new database to overcome any problems and input the data for the five facilities as a means of testing the new design. He hoped that this would demonstrate the viability of reformatting and importing all of the EPA data into the new design. He knew that the data would be used by the Institute for years to come and he was concerned that researchers would again and again need to deal with data anomalies for each analysis. Undoubtedly, assumptions would need to be made to resolve certain data ambiguities but at least they would be made explicit and uniformly applied to all future analyses.

7. REFERENCES

Center for Disease Control (1999). Public Heath Statement Mercury, Agency for Toxic Substances and Disease Registry. CAS#: 7439-97-6, March 1999.

EPA (2011a). Fact sheet: mercury and air toxics standards for power plants. Retrieved March 7, 2014 from http://www.epa.gov/mats/pdfs/20111221MATSsummaryfs.pdf

EPA (2011b). National emission standards for hazardous air pollutants from coal- and oil-fired electric utility steam generating units and standards of performance for fossil-fuel-fired electric utility, industrial-commercial-institutional, and small industrial-commercial-institutional steam generating units. 1 -1117. Retrieved March 7, 2014 from http://www.epa.gov/mats/pdfs/20111216MATSfinal.pdf

EPA (2013a). Overview – The Clean Air Act Amendments of 1990. Retrieved March 7, 2014 from http://epa.gov/air/caa/caaa_overview.html

EPA (2013b). 112(n) - Studies: Utilities, Coke Ovens, POTWs, Mercury, Hydrogen Sulfide. Retrieved March 7, 2014 from http://www.epa.gov/ttnatw01/112n/112npg.html

EPA (2013c). Overview by Section of CAA: Introduction to CAA and Section 112 (Air Toxics). Retrieved March 7, 2014 from http://www.epa.gov/ttnatw01/overview.html

EPA (1997). Mercury Study Report to Congress. *Executive Summary*. Retrieved March 7, 2014 from http://www.epa.gov/ttn/atw/112nmerc/volume1.pdf

EPA (2009). ICR Final Part A. Retrieved March 7, 2014 from http://www.epa.gov/ttn/atw/utility/g1/eu_mact_icr_part_a.pdf

Lemonick, M. D.(1990, November 5). Forecast: Clearer Skies: The Revised Clean Air Act is Costly but well worth the price. *Time*, *136(20)*,33.

Neville, A. (2012). Debate heats up over new mercury and air toxics rule. *Power Magazine.* Retrieved March 7, 2014 from http://www.powermag.com/debate-heats-up-over-new-mercury-and-air-toxics-rule/

Physicians for Social Responsibility. (2004). Mercury in Fish, Fact Sheet #3. Retrieved March 7, 2014 from http://action.psr.org/site/DocServer/Mercury_Fact_Sheet__1.pdf?docID=710

Note: Teaching Notes and Case Supplements are available by contacting the authors

Appendix

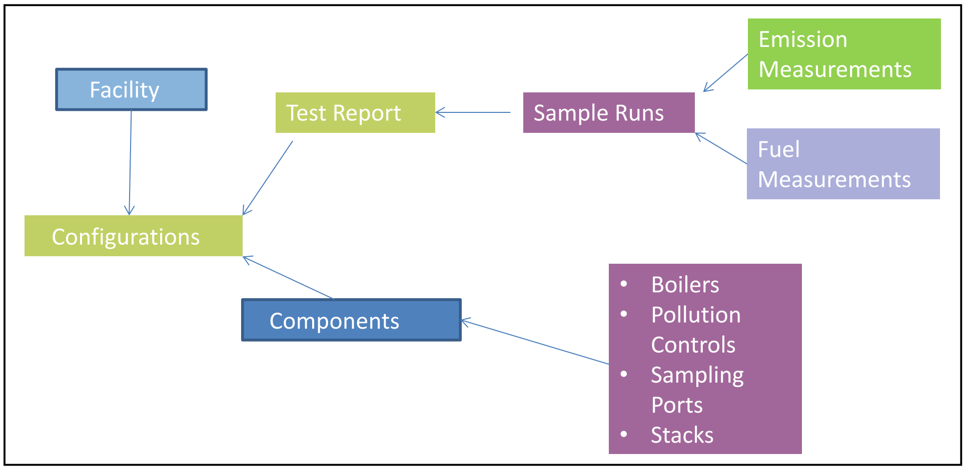

Figure 1 – Sketch of Entity Relationships

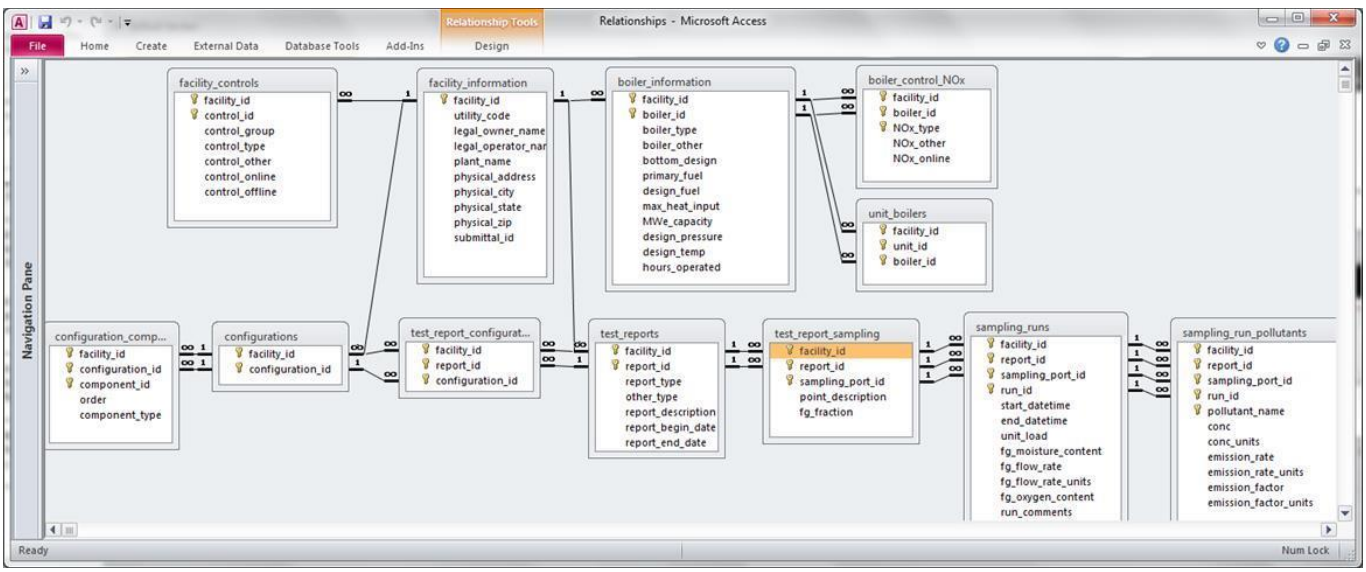

Figure 2 – M.S. Access Relationships in Test Database

Survey (see enclosure 1)	www.epa.gov/ttn/atw/utility/g1/eu_mact_icr_part_b.pdf
Data dictionary	www.epa.gov/ttn/atw/utility/pro/eu_mact_icr_part-i_ii-data_dictionary.pdf
Data Diagram	www.epa.gov/ttn/atw/utility/pro/eu_mact_icr_part-i_ii-db_erd.pdf
MS Access Database	www.epa.gov/ttn/atw/utility/eu_icr_parti_partii.mdb
Other Related Links	www.epa.gov/ttn/atw/utility/utilitypg.html

Table 1 - Links to Original Data and Descriptive Information

Potential Identifiers

1. **Facililty_ID**
2. **Configuration_ID**
3. **Boiler_ID**
4. **Unit_ID**
5. **Sampling_Port_ID**

Boiler characteristics

6. **Boiler_Firing_Type**
7. **Boiler_MaxHeatInput**
8. **MWe_Capacity**
9. **Primary_Fuel**

NOx Boiler Controls

10. **LoNox_Burner**
11. **Ovr_Fire** (Over air fire)
12. **Other_BoilerNOx**

NOx Facility Controls

13. **SCR** (selective catalytic reduction
14. **SNCR** (selective noncatalytic reduction)
15. **Other_Nox**

Mercury Facility Controls

16. **ACI** (activated carbon injection)
17. **DSI** (dry sorbent injection)

PM Facility Controls

18. **ESP** (Electrostatic precipitator)
19. **PM_Filter**
20. **PM_Scrubber**
21. **PM_Cyclone**
22. **PM_Other** (all other PM)

SO2 Facility Controls

23. **Wet_Fgd** (Wet Flue Gas Desulfurization)
24. **Dry_Fgd** (Dry Flue Gas Desulfurization)

Pollutant Emissions

25. **PM_F** (PM - Filterable)
26. **SO2** (Sulfur Dioxide - SO2)
27. **NOx** (Nitrogen Oxide - NOx)
28. **Hgt** (Total Mercury Hgt)
29. **HCl** – (Hydrogen Chloride HCl)

Table 2 – Structure of Extract

Boiler Firing Types

1. Tangential Firing
2. Wall Firing
    - 2.1. Front Wall Firing
    - 2.2. Rear Wall Firing
    - 2.3. Opposed Wall Firing
    - 2.4. Other Wall Firing
3. Cyclone Firing
4. Fluidized Bed Firing
5. Stoker Firing
    - 5.1. Stoker Underfeed
    - 5.2. Stoker Overfeed
    - 5.3. Stoker Spreader
    - 5.4. Stoker Other
6. Integrated Gas Combustion Cycle (IGCC)
7. Other Boiler Firing Type

Pollution Control Types

1. Particulate Matter (PM) Controls
    - 1.1. Electrostatic Precipitator (ESP)
        - 1.1.1. Cold Side ESP with Flue Gas Conditioning
        - 1.1.2. Cold Side ESP without Flue Gas Conditioning
        - 1.1.3. Hot Side ESP with Flue Gas Conditioning
        - 1.1.4. Hot Side ESP without Flue Gas Conditioning
    - 1.2. PM Filter
        - 1.2.1. Pulse Filter
        - 1.2.2. Reverse Air Filter
        - 1.2.3. Shake and Deflate Filter
    - 1.3. PM Scrubber
        - 1.3.1. Syngas
        - 1.3.2. Wet
        - 1.3.3. Venturi
    - 1.4. PM Cyclone
        - 1.4.1. Multiple Cyclone
        - 1.4.2. Single Cyclone
    - 1.5. PM other
2. Nitrous Oxide (NOx) Controls
    - 2.1. Facility Nox Controls
        - 2.1.1. Selective Catalytic Reduction
        - 2.1.2. Selective Non-Catalytic Reduction
        - 2.1.3. Facility Nox Other
    - 2.2. Boiler NOx Controls
        - 2.2.1. Boiler Nox Controls
        - 2.2.2. Low NOx Burner
        - 2.2.3. Overair fire (including advanced)
        - 2.2.4. Boiler NOx Other
3. Sulfur Dioxide(SO2) Controls
    - 3.1. Wet Flue Gas Desulfurization (WFGD)
        - 3.1.1. Wet FGD – Disk
        - 3.1.2. Wet FGD Flooded Disk
        - 3.1.3. Wet FGD Jet Bubbling Reactor
        - 3.1.4. Wet FGD Spray
        - 3.1.5. Wet FGD Tray
        - 3.1.6. Wet FGD Spray and Tray
        - 3.1.7. Wet FGD Other
    - 3.2. Dry Flue Gas Desulfurization (DFGD)
        - 3.2.1. Dry FGD Sorbent Injection
        - 3.2.2. Dry FGD Spray
        - 3.2.3. Dry FGD Other
4. Mercury Controls
    - 4.1. Activated Carbon Injection
    - 4.2. Dry Sorbent Injection
    - 4.3. Other Facility Controls
    - 4.4. Boiler Controls

Table 3 - Equipment Classification Hierarchy

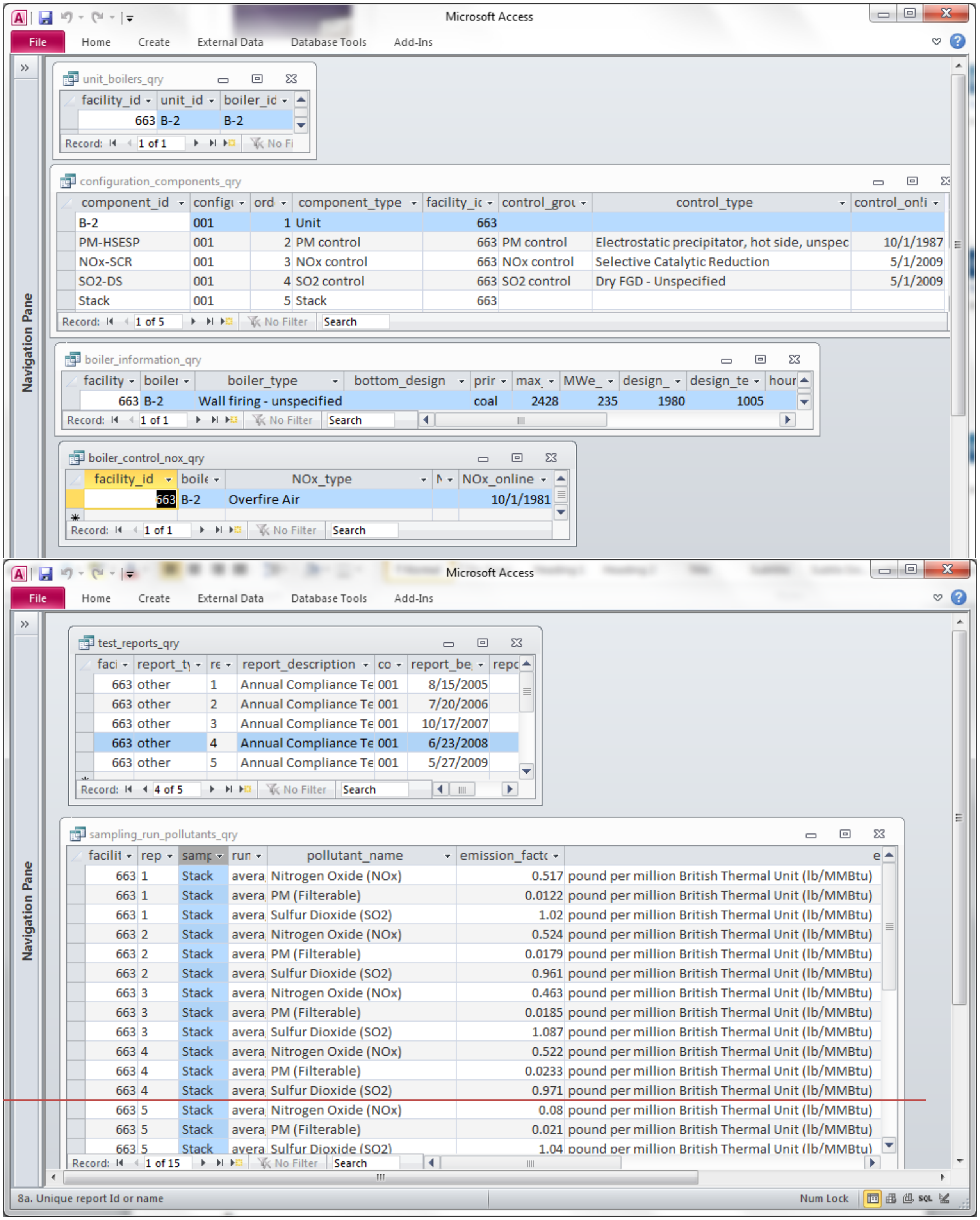

Figure 3 – MS Access Screen Shots for Facility 663

Extract Records[1]			
Potential Identifiers			
1. Facililty_ID	663	663	663
2. Configuration_ID	1	1	1a
3. Boiler_ID	B-2	B-2	B-2
4. Unit_ID	B-2	B-2	B-2
5. Sampling_Port_ID	Stack	Stack	Stack
Boiler characteristics			
6. Boiler_Firing_Type	Wall-firing	Wall-firing	Wall-firing
7. Boiler_MaxHeatInput	2428	2428	2428
8. MWe_Capacity	235	235	235
9. Primary_Fuel	coal	coal	coal
NOx Boiler Controls			
10. LoNox_Burner	0	0	0
11. Ovr_Fire (Over air fire)	1	1	1
12. Other_BoilerNOx	0	0	0
NOx Facility Controls			
13. SCR (selective catalytic reduction	0	0	1
14. SNCR (selective noncatalytic reduction)	0	0	0
15. Other_Nox	0	0	0
Mercury Facility Controls			
16. ACI (activated carbon injection)	0	0	0
17. DSI (dry sorbent injection)	0	0	0
PM Facility Controls			
1. ESP (Electrostatic precipitator)	0	1	1
2. PM_Filter	0	0	0
3. PM_Scrubber	0	0	0
4. PM_Cyclone	0	0	0
5. PM_Other (all other PM)	0	0	0
SO2 Facility Controls			
6. Wet_Fgd (Wet Flue Gas Desulfurization)	0	0	0
7. Dry_Fgd (Dry Flue Gas Desulfurization)	0	0	1
Pollutant Emissions[2]			
8. PM_F (PM - Filterable)	NULL	0.017975	0.021
9. SO2 (Sulfur Dioxide - SO2)	NULL	1.00975	1.04
10. NOx (Nitrogen Oxide - NOx)	NULL	0.5065	0.08
11. Hgt (Total Mercury Hgt)	NULL	NULL	NULL
12. HC1 – (Hydrogen Chloride HCl)	NULL	NULL	NULL
Applicable Dates[3]	10/81-9/87	10/87-4/09	5/09-present

1) Configurations shown here in columns would be in rows in the actual statistical extract.
2) The second configuration emissions are averages of first four reports 2005 - 2008. Only test report 5 emissions from 2009 should be associated with the third configuration.
3) 10/81 OverAir Fire Control installed. 10/87 ESP installed. 5/09 SCR and FGD installed.

Figure 4 Sample Extract Records for Facility 663

©2014 EDSIG (Education Special Interest Group of the AITP)