Paper Number LPS-3C

Establishing a Data Farm to Harvest Quality Information

Harold W Thomas (Hal)
Air Products and Chemicals, Inc.

Mike Moosemiller
Det Norske Veritas

Prepared For
35th Annual Loss Prevention Symposium
AIChE 2001 Spring National Meeting
Houston, TX
April 22-26 2001

November 2000
Unpublished

Abstract

To obtain useful reliability information from maintenance and inspection data once it has been collected requires effort. In the past, this has been referred to as data "mining", as if the data can be extracted in its desired form if only it can be found. In contrast, this paper proposes data "farming", and describes the seeds that are necessary to harvest the best possible crop of reliability information. The CCPS equipment reliability database project provides valuable lessons on how to "farm" rather than merely "mine" data. The CCPS work processes for establishing failure modes, populations to track, event data to collect, and implementation are all reviewed. Attention is given to knowing up front the data objectives and the quality of information desired. Also, the treatment of equipment surveillance periods turns out to be a critical variable for data quantity and quality; reasons for this and approaches to take are discussed. It will be seen that the quality and continuity of information that can be derived is much greater when the data sources can be "farmed" rather than "mined".

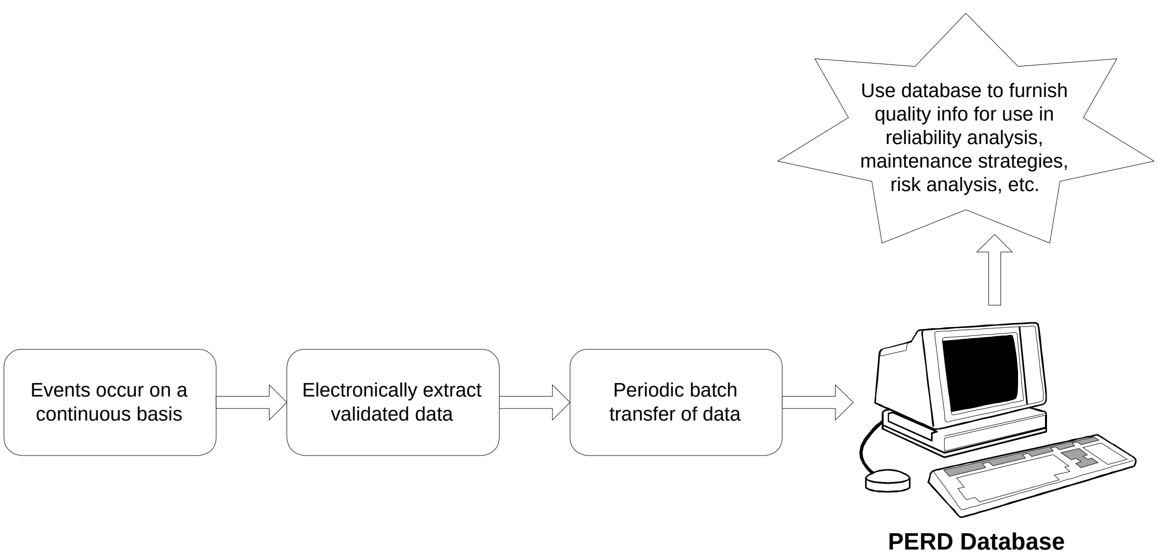
Introduction

The Center for Chemical Process Safety (CCPS), an AIChE industry technology alliance, has been facilitating an industry effort whose ultimate goal is to operate an equipment reliability database making available high quality, valid, and useful data to the HPI and CPI; enabling analyses to support availability, reliability, and equipment design improvements, maintenance strategies, support life cycle cost determinations; and provide better, more credible information for risk analyses. Along the way, participants learn what it takes to accomplish this goal and to take the knowledge developed during this effort and apply it to their management systems in a manner that adds value and lays the foundation for continued improvement.

Multiple companies from the oil, chemical, and industrial gas industries, as well as consultants, insurance companies, and equipment manufacturers have banded together to achieve this aim. These participants have come together under the aegis of CCPS which is a non profit organization dedicated to technical advancement and knowledge. CCPS is providing a forum to facilitate development and sharing of technical information with respect to an industry process equipment reliability database (PERD)[1].

This effort has built upon the ground breaking work published as part of prior initiatives such as IEEE-500[5], OREDA[7], and ISO standard 14224[6]. These efforts, however, have been limited to some extent with respect to the data quality that could be achieved because they relied to some degree on what this paper will refer to as data mining. When pursuing a data mining approach, data is extracted from whatever system exists and in whatever format and quality it exists. This forces multiple interpretations of raw data involving multiple persons in an attempt to create added value information. Accomplishing this type of exercise generally requires considerable manual labor each time data is extracted for analysis adding substantially to the cost. And due to inconsistencies during interpretation, the data quality is inherently suspect.

Figure 1



The Process Equipment Reliability Database (PERD) initiative seeks to improve the quality and lower the long term cost of data utilization by automating the process to the extent we can. We prefer to think of harvesting data as if we are operating a data farm. Figure 1 illustrates the concept. In chemical engineering vernacular, one can view it as a continuous batch process.

Success means that a fundamentally sound quality infrastructure is in place to support reliability analyses, maintenance strategies, risk analyses, and equipment improvement which in turn allows the expectation of continuous improvement and reliability growth[3]. This paper will review the key aspects of the CCPS work processes for establishing failure modes, the overall data structure, and implementation. Attention is given to knowing up front what the data objectives are and the quality of information desired. Also, the treatment of equipment surveillance periods is addressed as it turns out to be a critical variable for data quality.

Discussion of Key Reliability Terms

When establishing a fundamental foundation to build the operation, it is essential that the terms be defined. In the area of process plant and equipment reliability, this is especially true of the terms, failure, failure mode, failure cause, and failure mechanism. Attempts have been made to remain compatible with OREDA[7] and ISO 14224[6]. The true foundation of our use of these terms comes, however, from the paper, *The basic concepts of failure analysis*[8].

Equipment Failures and Related Circumstances

In the practical world, equipment and their components can "fail" under a variety of circumstances. There can be instant failures, partial failures, intermittent failures, and a number of other options. This invariably creates differing interpretations among engineers as to what happened, why it happened, and how it should be categorized, while at the same time wishing to be consistent in communication.

Similarly, the term "failure" is often confused with those of "fault" and "error."

For the purpose of this initiative, the collectively recognized term of **"failure mode"** has been adopted as a key term. As a description, it identifies how we "**observe a fault of an item,"** whether that fault is in its premature stage of failing, has been faulty for a period of time, or has subsequently resulted in the component failing to meet its expectation and has ultimately stopped performing. Shown below are excerpts of some key terms that begin to show their interrelationships.

Glossary of Terms - Excerpts

Failure	The termination of the ability of an item to perform a required function.
Failure cause	The circumstances during design, manufacture or use which have led to a failure.
Failure descriptor	The apparent, observed cause of failure.
Failure mechanism	The physical, chemical, or other process or combination of processes that has led to a failure.

Failure mode	The observed manner of failure. The failure modes describe the loss of required system function(s) that result from failures.

Practical Example - Failed Intercooler Tube of a Centrifugal Compressor

In this example, we will assume that there was a manufacturing defect in one of the intercooler tubes associated with a centrifugal compressor. This defect, in combination with the normal compressor vibration, resulted in the tube completely detaching from the tube sheet. The cooling water (tube side) circuit experienced an overpressure demand as the process operating pressure was greater than the cooling water circuit design pressure. The relief device in the cooling water circuit at the exchanger failed to adequately relieve the required capacity due to blockage in its outlet. Significant damage to the cooling tower resulted from the event. The compressor interlocks ultimately shut down the compressor, which caused a loss of product to the plant's customer.

In this example, if we start with the heat exchanger as an equipment system, PERD would define the failure mode as a "tube rupture." Its cause is due to a mechanical defect, and a contributing mechanism would be the vibration during operation.

From the compressor's perspective, the heat exchanger failure is a failure cause. As a result, multiple compressor failure modes occur, these being "Loss of process fluid to utility system" and "Fail while running." The failure mechanism for "Fail while running" would be the successful operation of the compressor interlocks. A complete list of potential centrifugal compressor failure modes as determined by a functional analysis is shown in Table 1. Table 2 lists possible incipient conditions. Capturing incipient condition information can provide valuable insight as to why or how actual failures occur. It is a mistake to refer to these as failure modes however, since they do not represent a loss of function.

Looking at the relief device, its failure mode is "Fail to relieve required capacity." The failure cause is blockage, and the mechanism is unknown as there is not enough information provided in the example.

Data Structure

In the previous compressor example, one could also analyze the effect on the plant as a whole. Its failure mode is "Fail while running" with an immediate cause of compressor shutdown. There are certainly other proximate causes associated with the heat exchanger, and if one were to take it to root cause, an understanding of what went wrong which allowed an exchanger with a manufacturing defect to be put into service would have to be considered.

Evaluating this single incident makes the importance of "engineering" the fundamental relationships into the data recording work process and tools used to capture the data. It is necessary to do this in a user-friendly manner, incorporating validation of data to the greatest extent practical, just as a farmer prepares the soil, making it easy to till and rich in nutrients, ready to yield a bumper crop.

If we are to capture data from incidents such as that illustrated by the compressor example, they must fit into an overall data structure such as that provided by the plant and equipment reliability database. Figure 2 shows the overall levels of the PERD taxonomy relationships, while Figure 3 illustrates the relationships in more detail at the specific equipment system level.

Table 1
Centrifugal Compressor Failure Mode

Complete Failures

- Fail while running
- Fail to start
- Fail to shut down
- Loss of process fluid to atmosphere
- Loss of process fluid to utility system
- Loss of utility fluid to atmosphere
- Loss of utility fluid to process system

Partial Failures

- Delayed Start
- Partial loss of throughput
- Leakage of process fluid to atmosphere
- Leakage of process fluid to utility system
- Leakage of utility fluid to atmosphere
- Leakage of utility fluid to process system
- Fugitive emission
- Surge
- Noise exceeds acceptable limit

Table 2
Centrifugal Compressor Incipient Conditions

Incipient Conditions

- High vibration
- Low suction pressure
- High discharge pressure
- Low discharge pressure
- High discharge temperature
- Low oil level
- High oil temperature
- Low oil pressure
- High oil filter delta pressure
- Low seal oil delta pressure
- Degraded oil condition
- High bearing temperature
- High motor amps

Figure 2
CCPS PERD Overall Taxonomy Data Structure

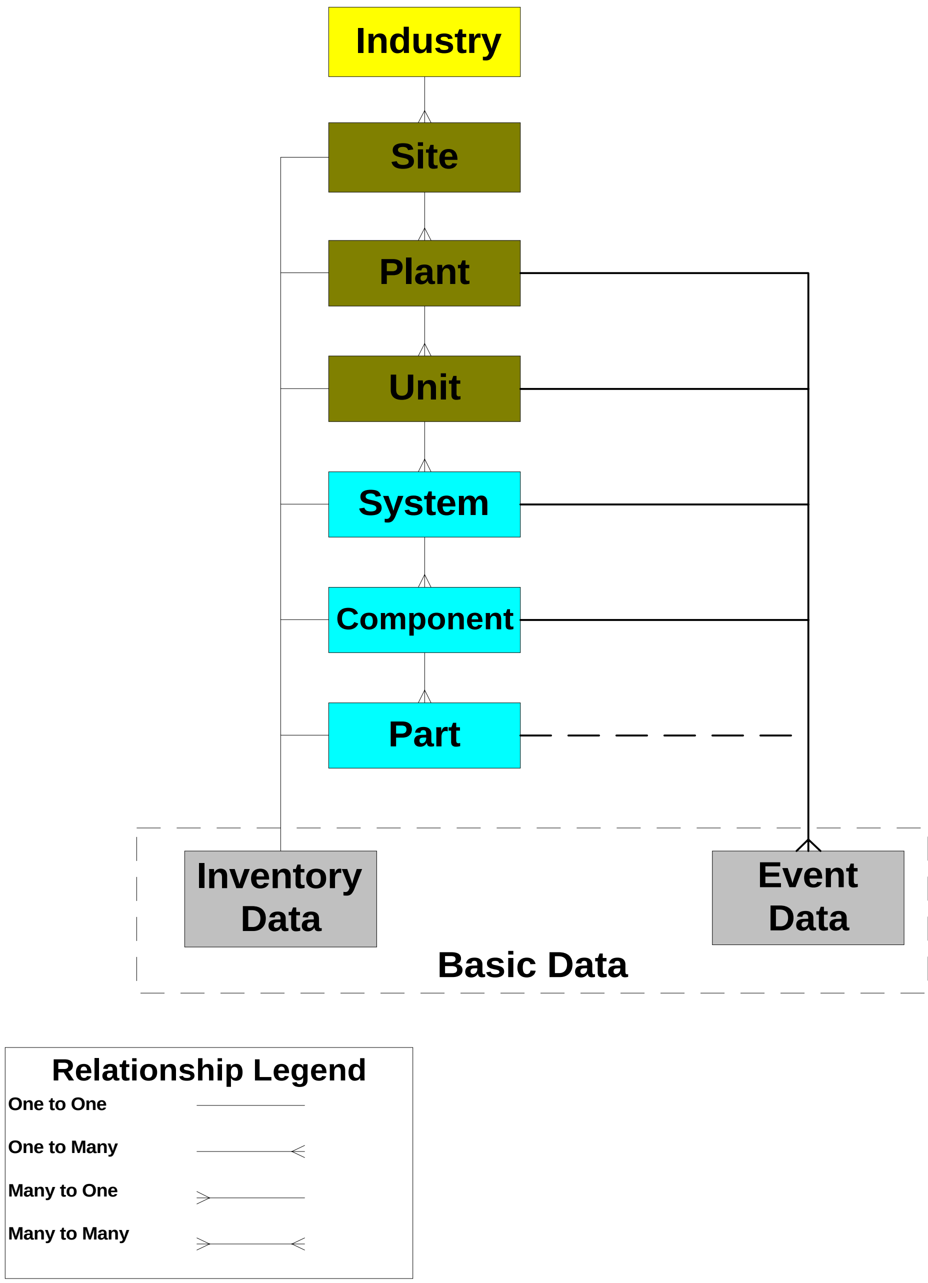
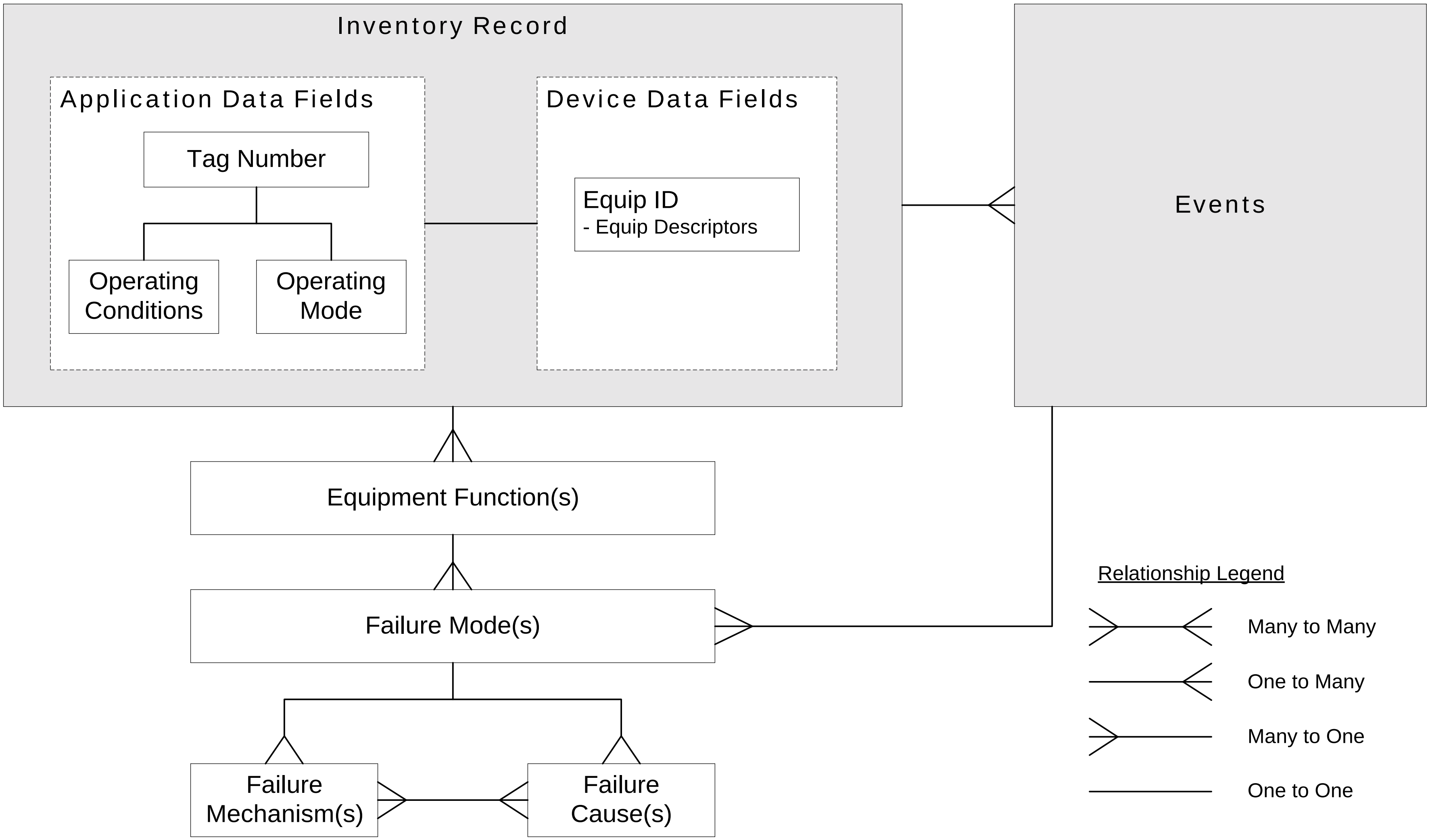


Figure 3
System Level Taxonomy Relationships



Taxonomy Development – An Engineering Foundation

The taxonomy development work process is at the core of the engineering foundation enabling the creation of a data farm. It provides the road map necessary to develop software tools that assist the harvesting of data, turning it into value added information along the way. Following the procedure establishes the:

- Equipment failure modes
- Data fields necessary to track application specific equipment populations
- Calculations that are necessary to establish reliability data
- Event data that if collected, infers specific failure modes

It provides excellent scope documentation in a consistent manner, regardless of equipment type, for the database software programmer to add new equipment taxonomies to the overall software application program. The basic procedure uses a rigorous step-by-step methodology. It provides an understanding of the role of the taxonomy and data field specifications in the database. The procedure takes advantage of basic OREDA[7] concepts and then uses the theoretical thought process expounded by Rausand and Oin[8], into a rigorous and practical methodology for creating equipment taxonomies.

The overall procedure for developing new equipment taxonomies is shown in Figure 4. The first step is to perform a functional analysis. The ultimate purpose of this is to determine the equipment's finite list of applicable failure modes. Prior to documenting the functions, it is necessary to develop the boundary diagram that explicitly shows what is included and what is excluded from consideration when a failure occurs. Figure 5 shows an example PERD boundary diagram for a remote actuated valve.

Table 3 displays the functions that were documented, that in turn determined the potential failure modes. The importance of identifying the complete list of fundamental failure modes cannot be understated. As indicated before, the CCPS PERD initiative owes a great deal to the work performed by OREDA as it served to form an initial foundation that inspired many of the PERD committee members. There was a conscious decision within the PERD initiative to emphasize quality and rigor prior to data collection, which is necessary to facilitate the ability to aggregate data harvested from disparate companies electronically, rather than using manual labor. Success in this endeavor offers the potential for greatly lowered cost of obtaining value added data, even as it improves the overall quality. The downside of going to the automated approach is that it lengthens the time to actual startup and initially lowers the quantity of what can be produced. To see what types of differences can arise by taking the methodical approach, an example has been provided in Table 4 comparing OREDA failure modes to proposed CCPS PERD failure modes for some level instrumentation.

Comparing the failure modes, one can note that some failure modes are not listed in the OREDA tables. It is the assumption of this paper that their work process only includes failure modes that were found to exist when mining existing data files. This is a reflection of the management system designs employed at the plants, their implementation, and the data mining approach. This is not to be construed as bad decisions by companies as they will invest in those things they believe add value. At this point, it has to be assumed that the data does not exist because it costs too much to obtain versus the value it provides. By identifying the complete list of fundamental failure modes, it enables companies to understand what they know and to also help them to realize what they do not know. If what they do not know is important, they have the beginning of a handle of what to do about it. Resolving and achieving a consensus of the total fundamental failure modes also provides an improved foundation for activities that attempt to define failure mode distributions, such as FMD-91[4].

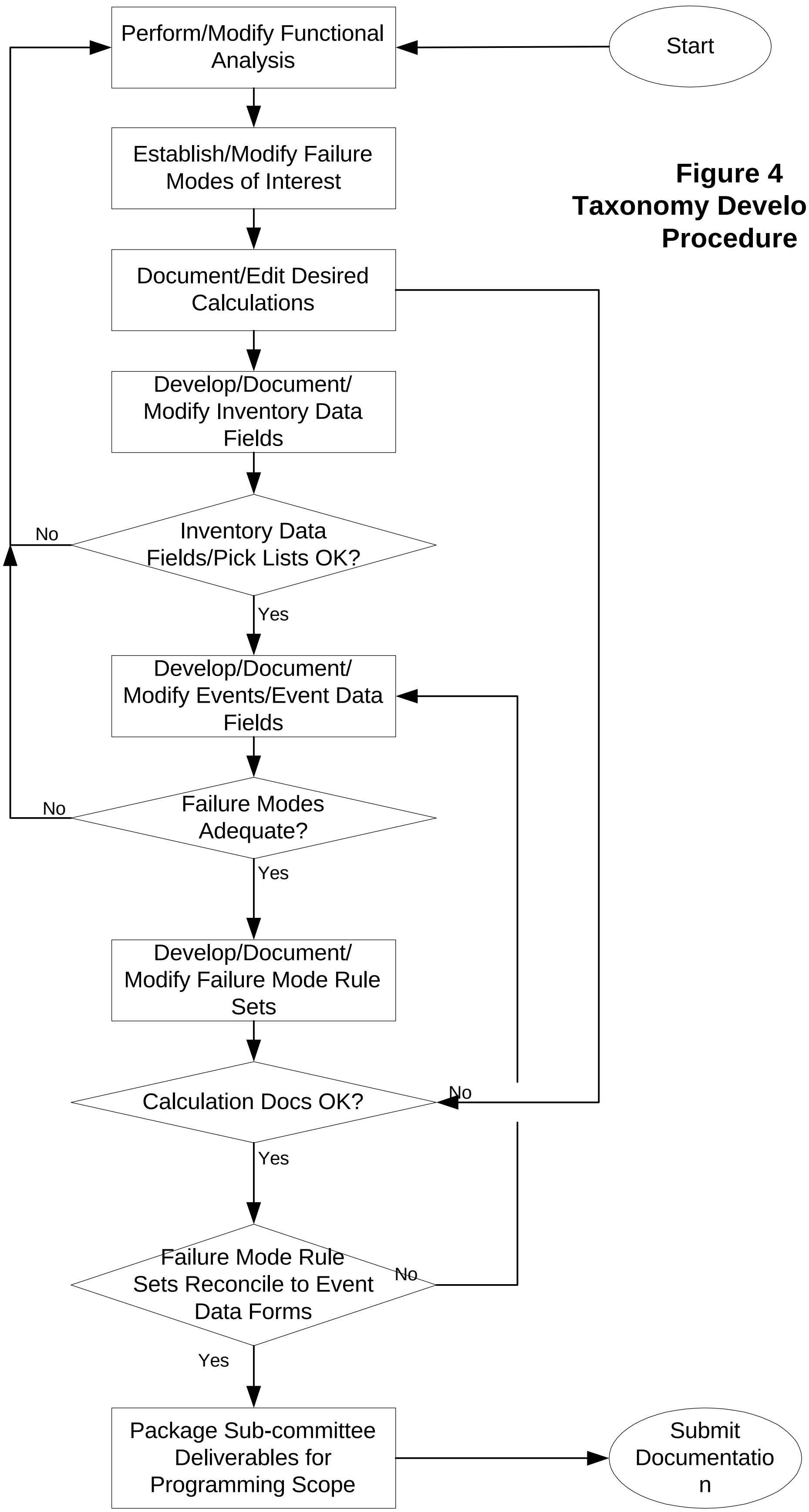


Figure 4
Taxonomy Development Procedure

Table 3
Remote Actuated Valve Functions and Failure Modes

Class	Function	Failure Mode	Failure Mode Description
Primary	Control of process variable	Frozen position	Valve position does not change with change of control signal to valve
		Spuriously opens	Valve goes to full open position even though control signal requires otherwise
		Spuriously closes	Valve goes to full closed position even though control signal requires otherwise
		Unstable control (hunting)	Valve unable to converge on stable control position
		Controlled variable high	Controlled variable stable at high value relative to control set point
		Controlled variable low	Controlled variable stable at low value relative to control set point
		Fail to close	Valve does not fully close upon demand.
		Fail to open	Valve remains fully closed upon demand
		Fail to hold position	Valve either opens or closes upon loss of utilities. Applies only to fail last position valves
Auxiliary	Maintain process fluid within pressure containing parts	Valve rupture	Body, bonnet, rupture
		Seal blow out	1)Blown packing 2)Blown body seal 3)Gasket Blowout
		External leakage	1)Packing leak 2)Body seal leak 3)Gasket leak 4)Body leak

Table 3 Continued

Class	Function	Failure Mode	Failure Mode Description
Protective	Provide positive shutoff within defined performance criteria	Seat leakage	Leakage in excess of defined performance criteria
	Limit maximum capacity	Excess flow	Valve capacity greater than capacity defined in design
	Provide adequate capacity when required for fail open over ride applications	Fail to open	• Valve remains fully closed upon demand OR • Valve fully plugged
		Reduced capacity	• Valve fails to completely open OR • Valve partially plugged
Information	Inference of flow relative to % travel	High indication of flow	Indication greater than actual but < 100%
		Low indication of flow	Indication less than actual but > zero %
	Position feedback	High indication	Indication greater than actual but < 100%
		Low indication	Indication less than actual but > zero %

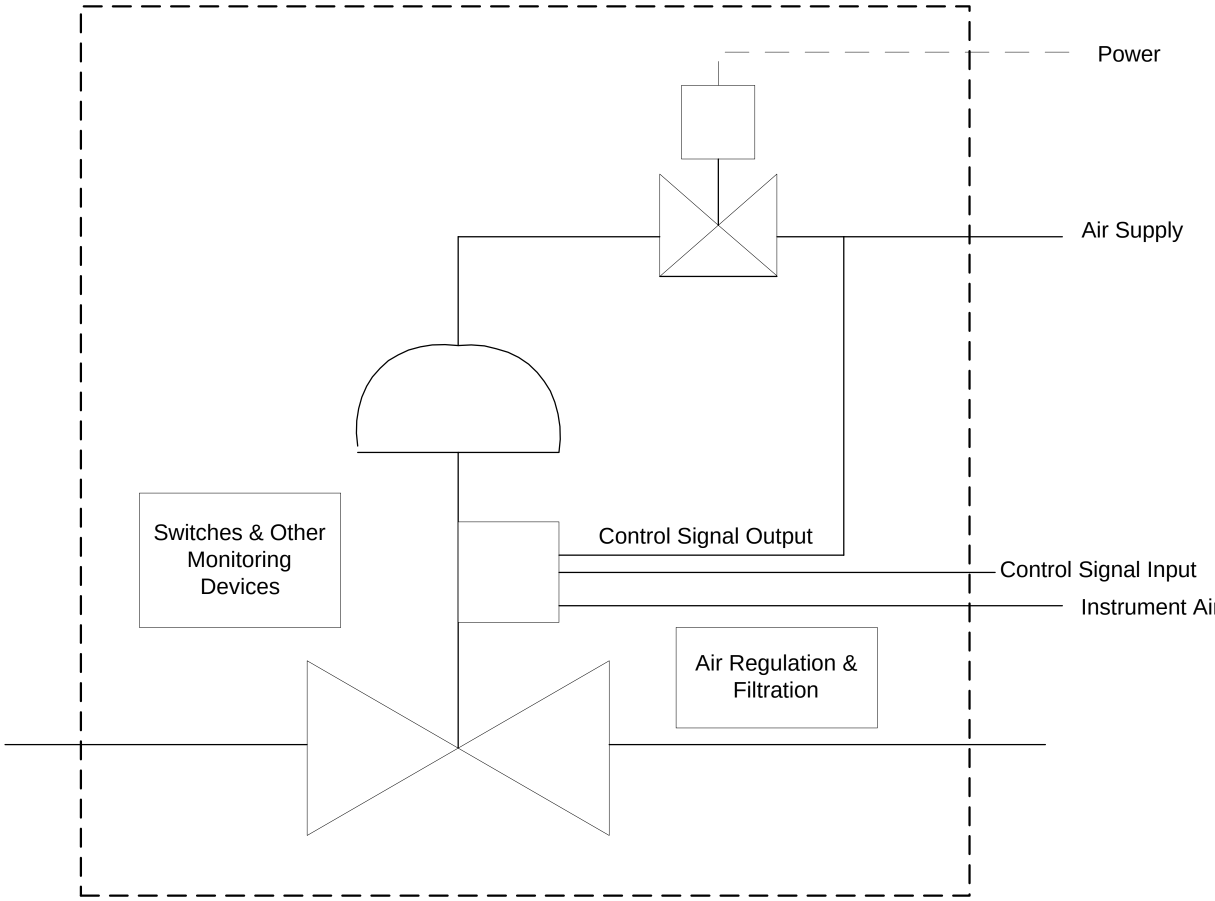
Table 4
CCPS VS OREDA Failure Modes

OREDA -92 Databook **Level Sensors - Transducer (1)**	**OREDA -97 Databook** **Process Sensor - Level (2)**	**CCPS** **Transmitter - Level (Includes Sensor)**	**CCPS** **Switch - Level (Includes Sensor)**
Critical Failure Modes	**Critical Failure Modes**	**Complete Failure Modes**	**Complete Failure Modes**
	• Fail to Operate on Demand		• Fail to Function on Demand
	• Spurious Operation		• Spuriously Functions
		• Output > 100 %	
• No Change of Output With Change of Input		• Frozen Output	
		• No Output	
• Unknown			
Degraded Failure Modes	**Degraded Failure Modes**	**Degraded Failure Modes**	**Degraded Failure Modes**
		• High Output	• Set Point High
• Low Output		• Low Ouput	• Set Point low
		• Output Slow to Respond	
		• Output Too Fast	
• Erratic Output	• Abnormal Output	• Output Erratic	
• Unknown	• Others		
Incipient	**Incipient**	**Incipient Conditions**	**Incipient Conditions**
• Contaminated • Unknown	• Minor in-service problems		
	Unknown		
	• Unknown		

Table 4 Notes:

1. Taxonomy consists of
   - Safety Systems
   - Process Alarm Systems
   - Level Sensors
   - Transducers (analog signals out, electric)
   - Application = Level measurement/alarm on process systems
   - Operational mode = Continuous
2. The boundary definition shown in OREDA-97 Figure 15, page 320 comprises the sensing element and the (local) electronics for signal conversion and transmission. The sensing element measures some process parameter (e.g. pressure, level, temperature, flow, etc,) and the electronics connects the measurement to a standard electric signal that is transmitted to a computer. Some sensors may be calibrated by adjusting a screw at the electronic housing. Isolation valves (block valves and associated pipe work are also included.

FIGURE 5 – Example Boundary Diagram – Remote Actuated Valve



It is next necessary to document calculations whose performance is expected following the availability of data. It is necessary to both define the statistical calculations that define each specific failure mode (i.e. fail while running, fail to start) as well as calculations that yield useful information applicable to the equipment system as a whole (i.e. time to restore).

Following this, the inventory data fields must be established. The general importance of inventory data is to allow specific sub populations to be analyzed. A small limited number of fields are necessary in order for the relational software to track the inventory as both a function of a unique equipment identification number and its specific installed location. Additional data fields are chosen after a standard industry specification has been identified to serve as a starting point basis. During the data field selection process, pick lists and other data validation techniques are determined and used to the greatest extent practical.

An important aspect of inventory data, are the specific failure modes derived from the original functional analysis. An example of this final documentation is shown in Table 5

Table 5

Surveillance Dates Tab		
Field Name	**Description**	
Failure Mode	**Complete Failures** • Fail to Start • Fail While Running • Fail to Shutdown • Loss of containment - 100% • Loss of purity **Partial Failures** • Delayed Start • Partial Shutdown • Production curtailment • Leakage • Spill • Release • Fugitive Emission	
Certified	One for each failure mode	yes/no
Surveillance Start Date	One for each failure mode	Date

Event handling within the PERD operation is very important. It is recognized that event data that yields potentially valuable information arises from a number of different circumstances. For instance, relief valve event data can arise from inspections, proof (function) testing, process demands, and other maintenance activities. It is necessary to determine what failure modes or other reliability data can be determined from each of the different event types. The question, "What data allows inference of specific desired failure modes and other reliability parameters?" must be answered. The answers to these questions directly determine what event data fields become part of the PERD taxonomy. It is required that the event data record only factual information that allows inference rules to be programmed so that the PERD software will consistently determine failure modes and provide validated data required for analysis.

Importance of Surveillance Start Stop Dates

It is important to recognize that no company records all the information required allowing determination of all potential failure modes. Typically, management systems exist that contain data only supportive of some failure modes that pertain to the equipment in question. As such, it becomes necessary to develop the application in a way that allows tracking of specific failure modes. In the quality plan that a data contributor prepares as part of the operating process, the inventory being tracked is documented. Included in this inventory are the specific failure modes that the companies management system is both capable of and committed to tracking. By doing this, if one of those tracked failure modes occur, it will be inferred upon entry of the applicable event data associated with the equipment item that failed. The quality plan documents how the management information systems will ensure this data will be recorded via the company work process and training.

Tracking individual failure modes makes it necessary to record surveillance start dates for each failure mode; the dates being a function of when the company has implemented a management system that ensures that when a failure mode occurs, the data that infers the failure mode as defined by the PERD taxonomy will be recorded. A key aspect of approving a company quality plan is the certification of the surveillance start dates.

Another aspect of the quality plan is the issue of surveillance stop dates which ensure that following transfer of one batch of quality data to the PERD database, the next batch will appear continuous and seamless with the prior submitted batch. The software provides for surveillance stop dates to be defined for all failure modes as part of downloading data in preparation for transfer to CCPS PERD. It is important however that the quality plan provide the documented work process to ensure good quality control and to support training of personnel.

Table 6 displays a typical report, accounting for individual surveillance start and stop dates. In the future, if a company chooses to improve its infrastructure and information management systems, additional failure modes may be included by formally revising the quality plan, allowing improved yields from the basic data crop. The taxonomy design and quality work process allows this to happen in a seamless fashion. This is a key aspect of the data farming concept, the ability to improve the crop and its yield over time as opposed to data mining, which simply has to accept the inherent quality of what exists.

The fact that the CCPS PERD initiative has documented the important technical information up front also allows companies to leverage that engineering when they choose to improve their systems. This can be as simple as improving the design of simple inspection forms or as substantial as helping the implementation of a new maintenance management system by lowering the engineering costs in certain areas and most importantly, helping ensure that promises made with respect to be able to use recorded data in a meaningful and cost effective fashion are met. Over time this not only improves the yields that these data farms can achieve, but vastly increases the acreage at bargain prices.

Operation Using the Quality Plan

Putting it all together requires a work process, the appropriate tools and training to achieve the concept initially shown in Figure 1. The overall work process, showing various lines of responsibility and how the different entities interface with one another is shown in Figure 6.

The initial Phase 1 software developed by Det Norske Veritas (DNV) has been completed, enabling the piloting activities to progress. Recognizing that the initial effort addressed a limited number of equipment types, sub committees have been formed, using the taxonomy development procedure to tackle additional equipment for inclusion in the tool.

Due to the complexity and the need to maintain a high level of quality assurance, it is necessary to make available a high quality training program to support the effort. GP, formerly General Physics, is rendering its services to make this happen. The training will initially help implement the work process and facilitate more complete testing of the software. As the process proceeds into full operation, it will be more important than ever to have adequate training resources in place to support its growth and to maintain the requisite level of quality.

Table 6

Average Failure Rate for Failure Mode

Report cutoff date: 28-Feb-2000
Run Date: 03-Nov-2000 Run Time: 09:42:31

Plants

Filtering Criteria

RESULTS OF ANALYSIS (time basis)

Confidence Limit: 90%

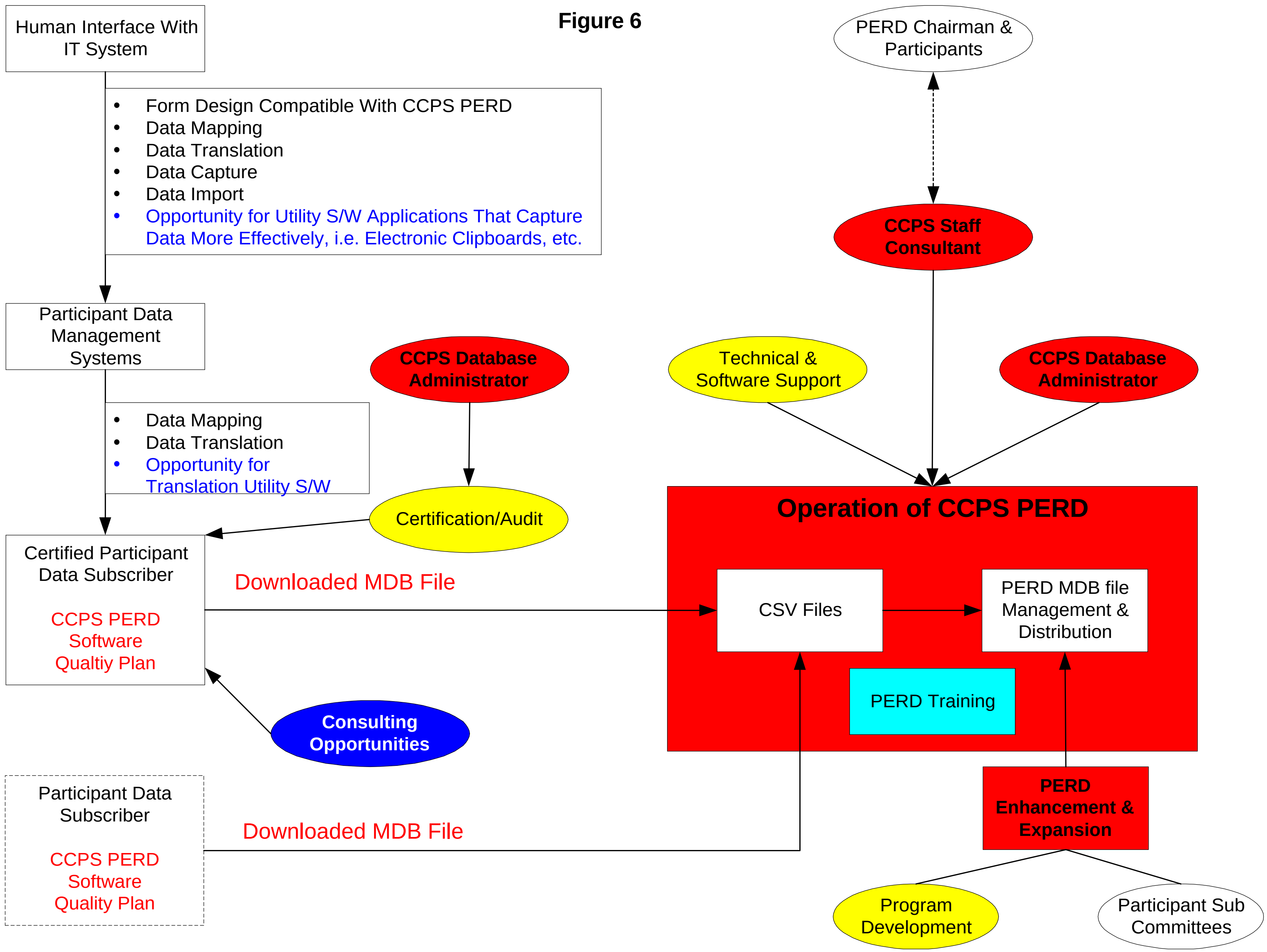
Population	Contrib Sites	Total Oper Time In Years	Failure Mode	No. of Failures	Lower (per year)	Mean (per year)	Upper (per year)
0	0	0.00	Fail to start	0	0.00	0.00	0.00
2	2	9.98	Fail while running	10	0.94	1.00	1.16
0	0	0.00	Fail to shutdown	0	0.00	0.00	0.00
0	0	0.00	Loss of containment 100%	0	0.00	0.00	0.00
1	1	3.59	Contaminated product	1	0.19	0.28	0.61
0	0	0.00	Delayed start	0	0.00	0.00	0.00
0	0	0.00	Partial shutdown	0	0.00	0.00	0.00
0	0	0.00	Curtailment	0	0.00	0.00	0.00
0	0	0.00	External leak	0	0.00	0.00	0.00
0	0	0.00	Spill	0	0.00	0.00	0.00
0	0	0.00	Release	0	0.00	0.00	0.00
0	0	0.00	Fugitive emission	0	0.00	0.00	0.00

report name: plant_failure_mode.rpt

Conclusion - The Data Farm

The CCPS PERD initiative is well on its way to establishing the basic rules and tools that will allow industry to engage in “data farming.” A work process has been established that allows a consistent approach to developing equipment taxonomies in a rigorous manner. Following this path is ultimately more cost effective than typical “data mining” exercises, once the work process has been implemented and personnel trained. Moreover, the potential data quality that can be harvested far surpasses anything that can be imagined today as the whole process is subject to defined quality control. By tracking individual failure modes and using the software programming to infer specific failures, users are in a position to understand what they know; in turn understanding what is not known, eliminating to a large extent assumptions as to failure mode distributions based more on gut feel than real science. If knowledge of a failure mode is important, the rules and tools are available to gain the necessary insight. Implementing the work process will go a long way towards enabling reliability growth in the CPI and HPI.

Figure 6


Human Interface With IT System
Form Design Compatible With CCPS PERD
Data Mapping
Data Translation
Data Capture
Data Import
Opportunity for Utility S/W Applications That Capture Data More Effectively, i.e. Electronic Clipboards, etc.
Participant Data Management Systems
Data Mapping
Data Translation
Opportunity for Translation Utility S/W
CCPS Database Administrator
Certification/Audit
Certified Participant Data Subscriber
CCPS PERD Software Qualtiy Plan
Downloaded MDB File
Consulting Opportunities
Participant Data Subscriber
CCPS PERD Software Quality Plan
Downloaded MDB File
PERD Chairman & Participants
CCPS Staff Consultant
Technical & Software Support
CCPS Database Administrator
Operation of CCPS PERD
CSV Files
PERD MDB file Management & Distribution
PERD Training
PERD Enhancement & Expansion
Program Development
Participant Sub Committees

Literature Cited

1. *CCPS Guidelines for Process Equipment Reliability Data*, AICHE, 1989
2. CCPS *Guidelines for Improving Plant Reliability Through Data Collection and Analysis*, 1998.
3. Criscimagna, Ned H. and Zsak, Michael G., *Reliability Growth,* The Journal of the RAC Reliability Analysis Center, Third Quarter 1998, pp. 1-5.
4. FMD-91, *Failure Mode/Mechanism Distributions*, Reliability Analysis Center, A DoD Analysis Center 1991.
5. IEEE Std 500-1984, *IEEE Guide to the Collection and Presentation of Electrical, Electronic, Sensing Component, and mechanical Equipment Reliability Data for Nuclear-Power Generating Stations*.
6. ISO 14224, International Standard, *Petroleum and natural gas industries – Collection and exchange of reliability and maintenance data for equipment*, First Edition, 1999.
7. OREDA-1992, *Offshore Reliability Data*, DNV Technica, Høvik, Norway, 1992
8. Rausand M. & Øien, K., *The basic concepts of failure analysis, In Reliability Engineering and System Safety* 53 (1996), pp. 73-83.

