Article citation info:
GALAR D, GUSTAFSON A, TORMOS B, BERGES L. Maintenance Decision Making based on different types of data fusion. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 14 (2): 135–144.

Diego GALAR
Anna GUSTAFSON
Bernardo TORMOS
Luis BERGES

MAINTENANCE DECISION MAKING BASED ON DIFFERENT TYPES OF DATA FUSION

PODEJMOWANIE DECYZJI EKSPLOATACYJNYCH W OPARCIU O FUZJĘ RÓŻNEGO TYPU DANYCH

Over the last decade, system integration is applied more as it allows organizations to streamline business processes. A recent development in the asset engineering management is to leverage the investment already made in process control systems. This allows the operations, maintenance, and process control teams to monitor and determine new alarm level based on the physical condition data of the critical machines. Condition-based maintenance (CBM) is a maintenance philosophy based on this massive data collection, wherein equipment repair or replacement decisions depend on the current and projected future health of the equipment. Since, past research has been dominated by condition monitoring techniques for specific applications; the maintenance community lacks a generic CBM implementation method based on data mining of such vast amount of collected data. The methodology would be relevant across different domains. It is necessary to integrate Condition Monitoring (CM) data with management data from CMMS (Computer Maintenance Management Systems) which contains information, such as: component failures, failure information related data, servicing or repairs, and inventory control and so on. These systems are the core of traditional scheduled maintenance practices and rely on bulk observations from historical data to make modifications to regulated maintenance actions. The most obvious obstacle in the integration of CMMS, process and CM data is the disparate nature of the data types involved, and there have benn several attempts to remedy this problem. Although, there have been many recent efforts to collect and maintain large repositories of these types of data, there have been relatively few studies to identify the ways these to datasets could be related. This paper attempts to fulfill that need by proposing a combined data mining-based methodology for CBM considering CM data and Historical Maintenance Management data. It shows a system integration of physical and management data that also supports business intelligence and data mining where data sets can be combined in non-traditional ways.

Keywords: *data mining, RUL, data fusion, condition monitoring, CMMS.*

W ostatniej dekadzie coraz częściej stosuje się integrację systemów, która pozwala przedsiębiorstwom zwiększać wydajność procesów biznesowych. Nowością w zarządzaniu infrastrukturą techniczną jest zwiększanie efektywności już poczynionych inwestycji w systemy kontroli procesów. Pozwala to zespołom do spraw operacyjnych, utrzymania ruchu oraz kontroli procesów monitorować i ustalać nowe poziomy alarmowe na podstawie danych o stanie fizycznym maszyn krytycznych. Utrzymanie urządzeń zależne od ich bieżącego stanu technicznego (condition-based maintenance, CBM) to filozofia utrzymania ruchu opierająca się na tym masowym poborze danych, wedle której decyzje dotyczące naprawy lub wymiany sprzętu zależą od jego obecnego oraz przewidywanego przyszłego stanu technicznego. Ponieważ dotychczasowe badania były zdominowane przez problem technik monitorowania stanu dla konkretnych aplikacji, nie opracowano ogólnej metody wdrażania CBM opartej na eksploracji (data mining) owych olbrzymich ilości zebranych danych, która miałaby zastosowanie w różnych domenach. Konieczna jest integracja danych z monitorowania stanu (condition monitoring, CM) z danymi dotyczącymi zarządzania pochodzącymi ze skomputeryzowanych systemów zarządzania utrzymaniem ruchu (CMMS), które zawierają informacje na temat uszkodzeń elementów składowych, dane związane z uszkodzeniami, a także informacje dotyczące obsługi lub napraw czy sterowania zapasami. Systemy te stanowią podstawę tradycyjnych praktyk obsługi planowej, a zasadzają się na całościowych obserwacjach dokonywanych na podstawie danych eksploatacyjnych, które pozwalają modyfikować regulowane działania obsługowe. Najbardziej oczywistą przeszkodą w integracji danych CMMS, danych procesowych oraz danych z monitorowania stanu jest rozbieżność ich natury. Dotychczas podjęto jedynie kilka prób rozwiązania tego problemu. Chociaż ostatnio wiele wysiłku włożono w gromadzenie i utrzymanie dużych zasobów tego typu danych, istnieje stosunkowo niewiele badań na temat możliwych sposobów powiązania owych zestawów danych. W prezentowanej pracy poczyniono próbę wypełnienia tej luki proponując metodologię łączoną opartą na eksploracji danych dla celów CBM, która bierze pod uwagę dane z monitorowania stanu i eksploatacyjne dane z zarządzania ruchem. W pracy przedstawiono integrację systemową danych fizycznych i danych z zarządzania, która wspiera także analitykę biznesową (business intelligence) oraz eksplorację danych, gdzie zestawy danych można łączyć w sposób nietradycyjny.

Słowa kluczowe: eksploracja danych, pozostały okres użytkowania (RUL), fuzja danych, monitorowanie stanu, CMMS.

1. Introduction

Maintenance can be considered as an information processing system that produces vast amount of data. However data is not synonymous with information; but that data must be processed with data analytical tools to extract the information, [5]. IT (Information Technology) and AI (Artificial Intelligence) tools development support the unprecedented transformation from the industrial age to the information age in maintenance using these existing and emerging technologies that analyze near real-time assets systems data to provide prediction and response maintenance capability. Several technological advances and initiatives at various levels have made a move toward CBM (Condition Based Maintenance) a reality for today's industry.

The transition to CBM requires a collaborative effort on a massive scale and is contingent on identifying and incorporating enhanced and emerging technologies into existing and future production systems. This will require new tools, test equipment, and embedded on-board diagnosis systems. Even more critical, the transition to CBM involves the construction of data-centric, platform-operating capabilities built around carefully developed robust algorithms. This will allow maintenance personnel in the field, support analysts, and engineers the ability to simultaneously, and in real-time, translate conditional data and proactively respond to maintenance needs based on the actual condition

Nowadays, two main systems are implemented in most maintenance departments: Computer Maintenance Management Systems (CMMS) are the core of traditional maintenance record-keeping practices and often facilitate the usage of textual descriptions of faults and actions performed on an asset. Second one is condition monitoring systems; recently developed Condition Monitoring Systems (CM) are capable of directly monitoring asset components parameters; however, attempts to link observed CMMS events to CM sensor measurements have been fairly limited in their approach and scalability.

A CBM strategy, where the optimal time to schedule a service visit is forecasted based on the condition of the equipment, is often proposed as an answer to the challenge of increase the efficiency and reduces the cost for the service of their equipment over their lifecycle, [10]. However, predictive maintenance approaches are frequently hampered. First, by the lack of knowledge of the features those give a good indication of the condition of the equipment. Second, by the processing power needed for prediction algorithms to forecast the future evolution of the selected features, especially, when large measurements are collected. To overcome these problems, this paper proposes to use data mining to improve the quality of the prognosis. Therefore,

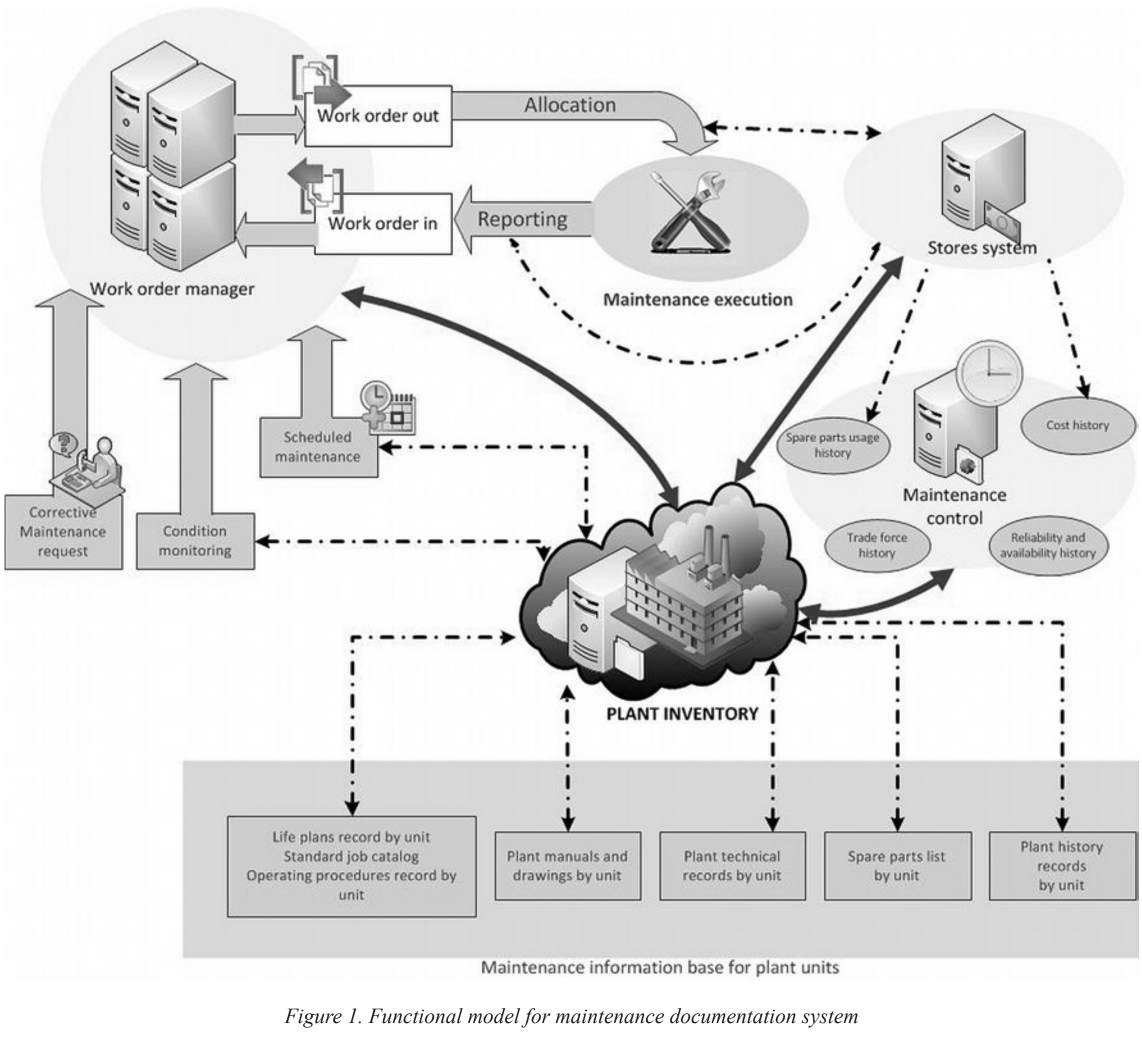


Figure 1. Functional model for maintenance documentation system

the development of future maintenance information systems in order to improve automatic condition monitoring systems enabled by embedded electronics and software in industrial machines, is one of the most important current research problems in this topic.

In this paper, the issues and challenge of this necessary integration of data of different nature are presented. It can be argued that understanding the requirements and constraints in conjunction - from maintenance AI and IT perspectives - is necessary to provide different decisions for different end users.

2. Maintenance historical data

2.1. Existing data in maintenance function

Maintenance documentation system, for recording and conveying information, is an essential operational requirement for all the elements of the maintenance management process. Maintenance documentation can be defined according to [11] as: Any record, catalog, manual, drawing or computer file containing information that might be required to facilitate maintenance work. Simultaneously, a maintenance information system can be defined as: The formal mechanism for collecting, storing, analyzing, interrogating and reporting maintenance information.

The way in which a maintenance documentation system generally functions is shown in Figure 1, a model which has evolved, according to [12], over a number of years through extensive studies of both paper-based and computerized systems, and which therefore illustrates the principal features of both types – features which, inevitably, they have in common. The system can be considered to be made up of the following interrelated modules:

1. Plant inventory,
2. Maintenance information base,
3. Maintenance schedule,
4. Condition monitoring,
5. Maintenance control.

The plant inventory (1) is a coded list of the plant units. This is the main way into the system. The asset data shall be collected in an organized and structured way. The major data categories for equipment are the following:

1. Classification data, e.g. industry, plant, location, system;
2. Equipment attributes, e.g. manufacturer's data, design characteristics;
3. Operation data, e.g. operating mode, operating power, environment.

These data categories shall be general for all equipment classes. Additionally some data specific for each equipment class (e.g. number of stages for a compressor) is needed. Finally, the classification of equipment into technical, operational, safety related and environmental parameters is the basis for the collection of assets data due to the different nature of devices (safety instrumented systems, productive assets, maintenance tools, condition monitoring systems etc.). This information is also necessary to determine if the data are suitable or valid for various applications. There are some data that are common to all equipment classes and some data that are specific for each equipment class.

These plant inventory units are the target for maintenance actions. These actions are basically of two kinds according to [3]:

- Those being done to correct an item after it has failed (corrective maintenance). It is required that for recording the reliability of an item, as a minimum corrective maintenance to correct a failure shall be recorded.
- Those being done to prevent an item from failing (preventive maintenance). A part of this may only be checks. Recording actual preventive maintenance (PM) is recommended to be done, essentially in the same way as for corrective actions. This may give additional information as follows:
  - the full lifetime story of an item (all failures and maintenance);
  - the total resources used on maintenance (man-hours, spare parts);
  - the total down time and hence, total equipment availability, both technical and operational;
  - the balance between preventive and predictive maintenance (inspections, tests) to verify the condition of the equipment to decide if any preventive maintenance is required or not.

Figure 2 shows the main types of maintenance actions being commonly performed.

According to [2] a new maintenance type called "opportunity maintenance" can be included if one considers maintenance of an item that is deferred or advanced in time when an unplanned opportunity becomes available.

Maintenance actions are the result of implemented maintenance program. Choice of applied methodology, ratio preventive-corrective, etc. is always up to maintenance manager who has plant inventory and maintenance information base to construct the model to be used. The maintenance information base (2) is a database of maintenance information, e.g. unit life plans, job catalog, etc. for each of the units. These data are characterised by:

- identification data; e.g. maintenance record number, related failure and/or equipment record;
- maintenance data; parameters characterising a maintenance, e.g. date of maintenance, maintenance category, maintenance activity, impact of maintenance, items maintained;

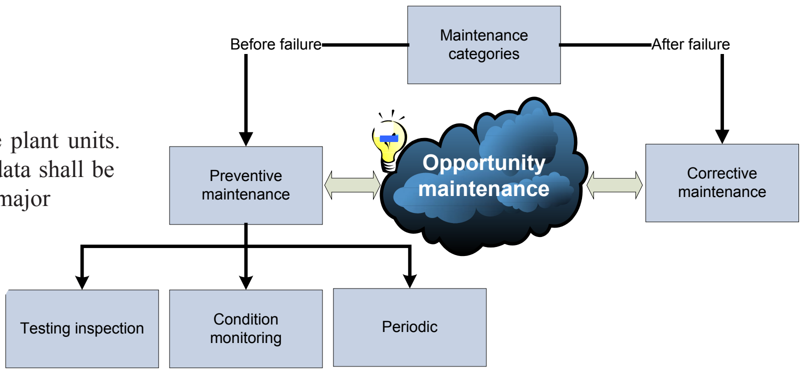
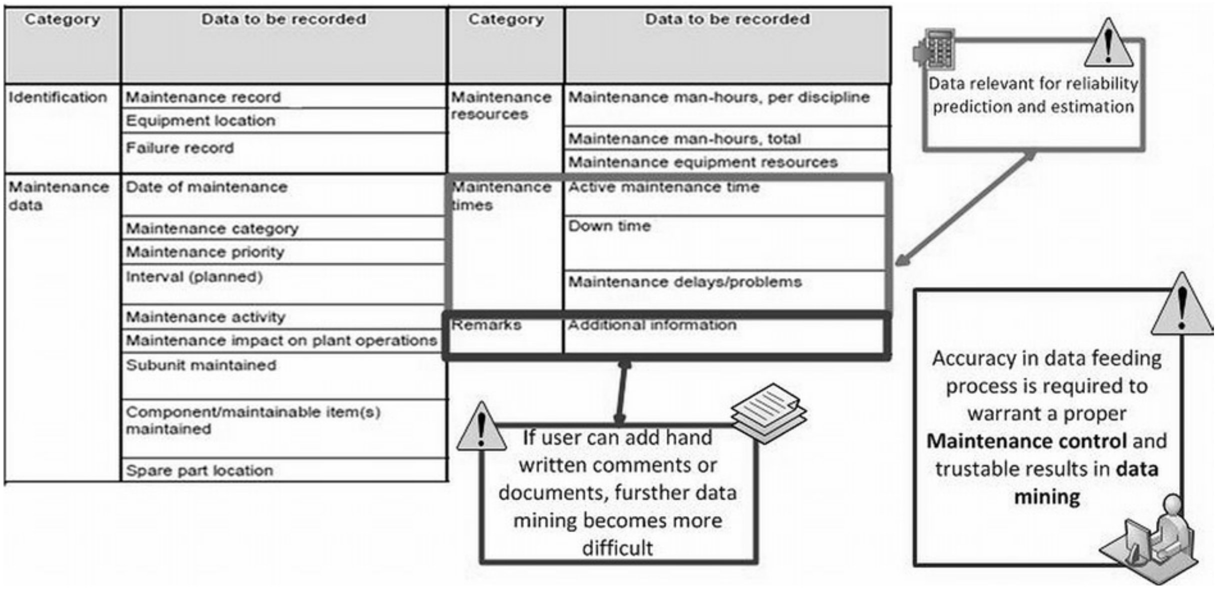


Figure 2. Maintenance actions categorisation



Category	Data to be recorded	Category	Data to be recorded
Identification	Maintenance record	Maintenance resources	Maintenance man-hours, per discipline
	Equipment location		Maintenance man-hours, total
	Failure record		Maintenance equipment resources
Maintenance data	Date of maintenance	Maintenance times	Active maintenance time
	Maintenance category		Down time
	Maintenance priority		Maintenance delays/problems
	Interval (planned)	Remarks	Additional information
	Maintenance activity		
	Maintenance impact on plant operations		
	Subunit maintained		
	Component/maintainable item(s) maintained		
	Spare part location		

Table 1. Maintenance records recommended meeting most International Standards and general recommendations [9]

- maintenance resources; maintenance man-hours per discipline and total, utility equipment /resources applied;
- maintenance times; active maintenance time, down time.

A common report for all equipment classes should be used for reporting maintenance data. The data required are shown in Table 1. For some equipment classes minor adaptations may be needed. The minimum data needed to meet the objectives of International Standards, Maintenance Association and CMMS manufacturer's recommendations are identified in Table 1.

Recording maintenance actions is crucial for a successful further knowledge extraction that is why all actions performed should be recorded. PM records are mainly useful for the maintenance engineer, but will also be useful for the maintenance engineer wanting to record -or estimate- the availability of equipment, and doing lifetime analysis not only taking failures into account, but also maintenance actions intended to restore the item to "as-good-as-new" condition. PMs are often performed on a higher indenture level (e.g. "package" level); hence there may not be data available that can be related to the items on the lower indenture level. This restriction must be considered when defining, reporting and analysing PM data.

During the execution of PM actions, impending failures may be discovered and corrected as part of the PM activities. In this case the failure(s) shall be recorded as any other failure with the subsequent corrective action done even though it initially was considered to be a PM type activity. The failure detection method shall in this case be referred to as the type of PM being done. It is, however, realised that some failures, generally of minor character, may be corrected as part of the PM and not recorded specifically. The practice on this may vary between companies and should be addressed by the data collector(s) in order to reveal the possible type and amount of failures being included within the PM program.

A final option is to record the planned PM program as well. In this case it is possible to additionally record the differences between the planned and the actual performed PM (backlog), [4]. An increasing backlog will be an indication that the control of the conditions of the plant is being jeopardised and may in adverse circumstances lead to equipment damage, pollution or personnel injury.

Regarding corrective maintenance, failure records are especially relevant for further knowledge extraction so failure data have to be recorded in a proper way to be suitable for further computation. A uniform definition of failure and a method of classifying failures are essential when data from different sources (plants and operators) need to be combined in a common maintenance database.

These failure data are characterised by:

- identification data; e.g. failure record number and related equipment that has failed;
- failure data for characterizing a failure, e.g. failure date, items failed, failure impact, failure mode, failure cause, failure detection method, so on.

The type of failure and maintenance data shall normally be common for all equipment classes, with exceptions where specific data types need to be collected. Corrective maintenance events shall be recorded in order to describe the corrective action following a failure.

Finally the combination of plant inventory and maintenance base information produces the expected maintenance schedule. This schedule is a mixture of available techniques to fulfill stakeholder's constraints and achieve company goals. This mixture is usually composed by some scheduled maintenance and condition monitoring to perform condition based maintenance. The maintenance schedule (3) is a schedule of the preventive maintenance jobs (over a year and longer) listed against each of the units in the life plans. The condition monitoring schedule (4) is a schedule of the condition monitoring tasks, e.g. vibration monitoring listed against each of the units in the life plans. Preventive maintenance records are required to retain the complete lifetime history of an equipment unit.

The system has to plan and schedule preventive jobs (arising from the maintenance schedule), corrective jobs (of all priorities) and where necessary modification jobs. The jobs are carried out by trade-force via hard copy or electronic work orders. Information coming back on the work orders (and other documents) is used to update the planning systems and provides information for maintenance control. The maintenance control system (5) uses information coming from a number of sources, work orders, stores, shift record, etc. to provide various reports for cost control, plant reliability control, etc.

One of the main issues is the integration of these data with the rest of company records such us health and safety, finances, etc.. Up until about 10 years ago most CMMS were stand alone, i.e. they had no electronic linkage with other company software. The most recent computerized maintenance systems are integrated electronically (they are in the same database) with stores, purchasing, invoicing, company costing, payroll and also can have electronic links to project management and condition monitoring software.

2.2. The search of a comprehensive data format

Each mentioned data becomes a database record, e.g. a failure event, shall be identified in the database by a number of attributes. Each attribute describes one piece of information, e.g. the failure mode. It is recommended that each piece of information be coded where possible. The advantages of this approach versus free text are:

- facilitation of queries and analysis of data;
- ease of data input;
- consistency check undertaken at input, by having pre-defined code-lists;
- minimise database size and response time of queries.

The range of pre-defined codes should be optimised. A short range of codes will be too general to be useful. A long range of codes will give a more precise description, but will slow the input process and may not be used fully by the data collector. Selected codes shall, if possible, be mutually exclusive. The disadvantage of a pre-defined list of codes versus free text is that some detailed information may be lost. It is recommended that free text is included to provide supplementary information. A free-text field with additional information is also useful for quality control of data. This free text box is extremely risky in further data mining process due to difficulties of text recognition and interpretation, see Table 1. Different employees have different skills to describe failures, events and actions and expert systems are not so good to distinguish all these variations.

For all mentioned categories, it is recommended to include some additional free text giving more explanatory information as available and deemed relevant, e.g. include a more verbal description of the occurrence leading to a failure event. This would assist in quality checking the information and browsing through single records to extract more detailed information. However users should be aware of the existing risk in automatic processing of these records.

2.3 Database structure

The data collected shall be organised and linked in a database to provide easy access for updates, queries and analysis. Several commercial databases are available that can be used as main building block for designing a reliability database. Two aspects on organising the structure of data shall be addressed as follows:

- Logical structure: This defines the logical link between the main data categories in the database. This model represents an application-oriented view of the database. The example in Figure 3 shows a hierarchical structure with failure and maintenance records linked to the classification/equipment description (inventory). Records describing preventive maintenance (PM) are linked to the inventory description in a many-to-one relation. The same applies for failures, which additionally have related corrective maintenance records linked to each failure. Each record (e.g. failure) may consist of several attributes (e.g. failure date, failure mode, etc.).
- Database architecture: This defines the design of the database as to how the individual data-elements are linked and addressed. Four model categories are commonly available ranked in order of complexity and versatility:
  - Hierarchical model: Data fields within records are related by a 'family tree-like' relationship. Each level represents a particular attribute of data;
  - Network model: This is similar to the hierarchical model; however, each attribute can have more than one parent;
  - Relational model: The model is constructed from tables of data elements, which are called relations. No access path is being defined beforehand; all manipula-

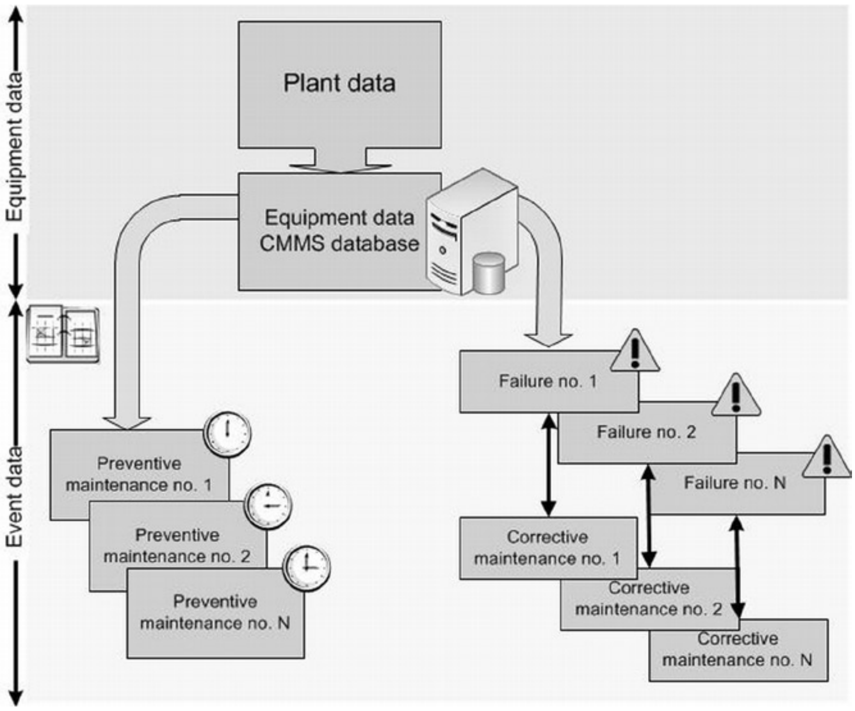


Figure 3. Logical data structure

tion of data in tabular form is possible. The majority of database designs use this concept;
- Object model: Software is considered as a collection of objects that each has a structure and an interface. The structure is fixed within each object while the interface is the visible part that provides the link address between the objects. Object modelling enables the database design to be very flexible, extendable, reusable and easy to maintain. This model seems to be the most popular in new database concepts.

3. Condition Monitoring data and automatic asset data collection

Condition monitoring involves comparing on-line or off-line data with expected values; if necessary, it should be able to generate alerts based on preset operational limits. Health assessment determines if the health of the monitored component or system has degraded, and conducts fault diagnostics. The primary tasks of prognostics involve calculating the future health and estimating the remaining useful life (RUL). In reality, however reliable and effective CBM faces some challenges. First, initiating CBM is costly. Often the cost of instrumentation can be quite large, especially if the goal is to monitor equipment that is already installed. It is therefore important to decide whether the equipment is important enough.

Implementing condition-based maintenance requires the setting of an information system to meet the basic requirements of:

- Collection and processing of large quantity of information not previously available, regarding the condition of each part of a machine.
- Initiate corrective maintenance actions within the lead-time (the period of time between the off-limits condition and an emergency shutdown). In this respect there may be two different situations which the examiner may encounter :
  - The condition of machine is not yet close to breakdown. In this case the normal procedure through the maintenance planning section will be followed.
  - The condition of machine is already well within the lead-time (near to breakdown). In this situation the information must be directly passed on to the maintenance supervision for carrying out emergency corrective maintenance actions.

In order to operate the condition based maintenance program correctly, the maintenance personnel should introduce into the system:

- Condition of machine,
- Part of machine probably defective,
- Probable defect,
- Time during which failure must be repaired.

By scrutinizing and correlating of diagnosis against actual findings during repair work, it will be possible:

- To control the examiner training,
- To improve the correlation between parameters chosen for condition measurement and actual defects found,
- To obtain severity curves specific to each machine.

Making the potential of condition monitoring a reality requires that large amounts of data be collected, monitored, filtered and turned into actionable information. The cheaper and more ubiquitous the computerized monitoring hardware becomes, the greater the volume of data and the more challenging it becomes to manage and interpret. The vast amount of diagnostic data produced by today's smart field devices can be a very important source for accurate documentation of maintenance activities. But the sheer volume and complexity of such information can be daunting and difficult for maintenance personnel to manage. What's needed is an effective means of compiling and organizing the data for day-today utilization by your staff, while preserving and recording significant events for future reference. Data is becoming more and more available.

However, in most cases, this data may not be used due to its bad quality, or even properly stored for several reasons, [13]:

- Project managers do not have sufficient time to analyse the computerised data so they don't care about proper storage;
- The complexity of the data analysis process is beyond the capabilities of the relatively simple maintenance systems commonly used;
- There has been no well defined automated mechanism to extract, pre-process and analyse the data and summarise the results so stored data are not reliable.

Maintenance personnel, not only cope with large amounts of field-generated data, they turn that information to their advantage in a number of ways. Real Time Condition Monitoring (RTCM) systems produce lots of warnings, alarms and reports that can be used by maintenance people for many purposes. In this way, the most important issues are identified and handled quickly.

Ultimate goal is to fully integrate RTCM data with CMMS to generate work orders as needed. That will provide true automation from the time a field device begins to show signs of reduced performance until a work order is printed out in the maintenance department and a technician is dispatched to the scene. In Figure 4, this automation of work order dispatching is shown.

This level of integration of CMMS and CM is feasible due to IT evolution. With the development of open communication protocols, the information accumulated by smart field devices can be captured by asset management software. It's no longer necessary for technicians to carry handheld communicators or laptops into the plant to evaluate the condition of instruments, some of which are quite inaccessible or in hazardous areas, to be followed by manually documenting test results and current device status.

Current applications compile databases of every smart instrument used for process control, including its design parameters, original configuration, maintenance history and current operating condition. With these online tools, technicians can obtain up-to-date information on any device and they never have to make manual entries back into a system. Every event is recognized and recorded, whether initiated by a technician or caused by an external force such as an equipment breakdown or power failure. This process produces one immediate result for shop floor level because work orders can be open and closed helped by devices that collect automatically information and send warning if something wrong happens. Users can refer to recorded alerts to identify any devices that have been problematic over time and what corrective steps may have been taken previously. Automated documentation provides a seamless record of events in a given production area, including communication failures, device malfunctions and process variables

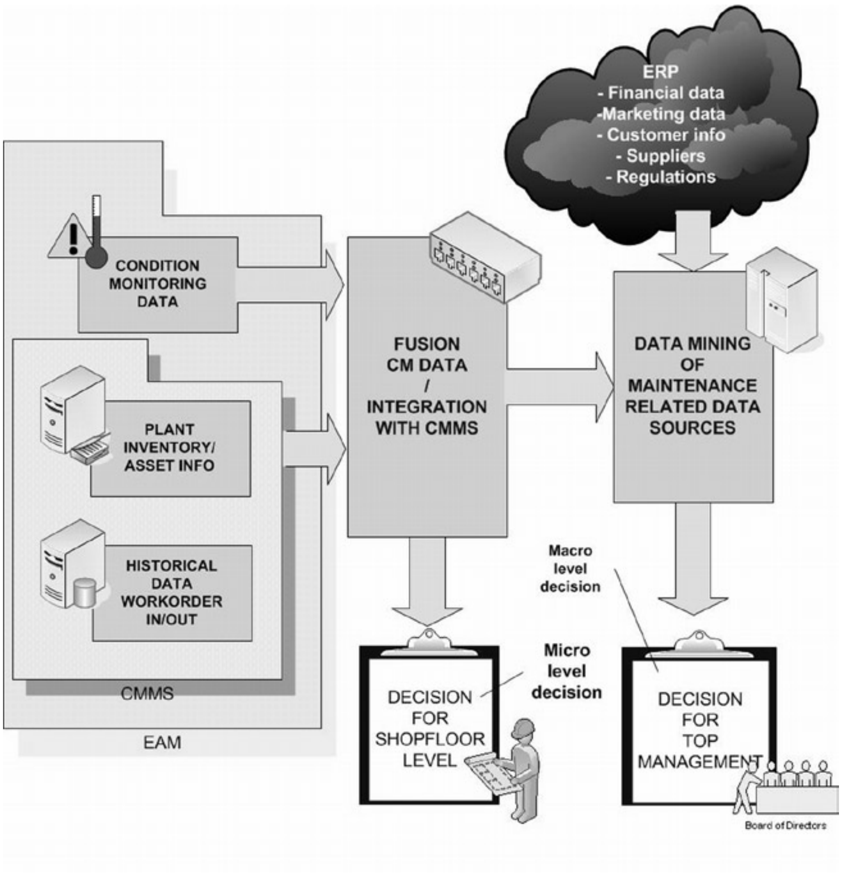


Figure 4.Two step integration of RTCM and CMMS databases

that are out of range. Armed with this information, maintenance personnel are better equipped to understand and resolve nagging repetitive issues to improve the process. If there is an issue, or if maintenance personnel are experiencing a rash of issues, they can go back into the records and get a sense of what's been going on over time. They can search by a specific device or by location.

Since all records are date and time stamped, users can easily determine when and by whom a particular device was changed or tested, including "as found/as left" notations. With this information in a database that cannot be edited, it should never be necessary for technicians to spend time searching for historical information on a device. Since events can also be recorded manually, users can document unusual occurrences affecting the entire plant, such as a lightning strike or power outage, or individual events like device inspections.

This decision level is extremely useful for technicians to take immediate actions. However vast amount of available information can produce new knowledge if it's exploited with proper AI tools due to real physical integration in same database types and locations. Modern CMMS information is stored in very large relational, or tabular, databases. This format is appropriate for an integration investigation since there are a large number of software tools available to query and investigate the tables. For the historical analysis, only certain fields are required, thus allowing for the previously mentioned sensitive data to be removed or filtered. The data subset still contains a full history of component faults and related actions, providing a comprehensive maintenance history profile while alleviating security concerns.

Importing CM data into this relational database is somewhat more challenging but possible, since each type of sensor generates different data classes, sampling rates, and number of compiled indicators. Furthermore, each manufacturer stores the collected information in unique proprietary formats, requiring platform-specific importation software to be written. However most CM software allows the CM data to be exported from the original interface so that it can be expanded and generalized.

Although both the CMMS and the CM data now co-exist within a single database where it can be queried and explored, automating the discovery of linked events requires additional processing. Relating a given maintenance fault or action, which is textual, to sensor data, which is some arbitrary data class type, can only be accomplished through the compilation of overlapping metadata, [14]. The fields which are generated characterize the location and significance of events, creating a quantified set of parameters by which the disparate data can be compared. Metadata for CM records is generated differently depending on the data class involved. One-dimensional and dimensionless quantities can be assigned rarity parameters through statistical distribution analysis, and higher dimensional data requires using neural networks to identify anomalies. Determining rarity is often accomplished through simple single variable statistical analysis, while severity is typically derived from developers recommended threshold values. More complex domain types require more advanced, though typically well-understood analyses such as neural networks which can isolate anomalous points from multidimensional data. It is predicted that through the integration process, more advanced metrics and indicators can be discovered which implement previously unexplored relationships in the data, such as multi-parameter trending. This new discovered knowledge can help maintenance personnel to find out the Remaining Useful Life of the system in order to schedule operation and maintenance processes in function of such relevant information. This information affects replacement of assets, shutdown of the plant, overhauls etc.; so it constitutes the second decision level displayed in Figure 4 which is strongly related to business goals and useless for immediate interventions.

4. Data mining of maintenance historical data for RUL calculation

Condition based maintenance (CBM) is the real target of all maintenance plans to optimize inspections and intervention avoiding wastes of time and money. It has evident benefits, including reducing maintenance cost, increasing machine reliability and operation safety, and improving time and resources management [1]. That is why major improvements can be achieved in maintenance cost, unscheduled machine failures, repair downtime, spare parts inventory, and both direct and indirect overtime premiums, by using CBM. Although these benefits are well illustrated, two major problems hamper the implementation of predictive maintenance in industrial applications. First, the lack of knowledge about the right features to be monitored and second, the required processing power for predicting the future evolution of features, which is often not available.

Data Mining (also known as Knowledge Discovery in Databases, or KDD) has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information, from data" [6]. Data mining and knowledge discovery can be applied to historical data from the field in order to optimally identify these relevant features for the condition of the equipment and the associated thresholds and contexts. Based on this information, a prediction model is fitted to the live data of the equipment, collected from customer's premises, for predicting the future evolution of these features and forecasting the RUL (remaining useful life) and consequently the time interval to the next maintenance action. To overcome the limited processing power of the machine's processor and data replications costs derived from the huge amount of collected and stored data, the concept of cloud computing comes up, performing prediction through computation of remote located databases.

The lack of knowledge for the proper extraction of right features can be compensated with data mining techniques proved to be useful for relevant features extraction. It has been proved in [8], [15] and [16] that application of data mining techniques to condition monitoring data is very useful for extracting relevant features which can be used as parameters for machine diagnosis and/or prognostics. However, in many other industrial applications, no clear physical understanding of the process is available and therefore to retrieve these relevant features, a clear methodology is required.

This paper proposes the use of data mining for CBM optimization in order to perform an accurate predictive maintenance scheduling. So far, prognostic has been tackled as an independent step from data mining, assuming the prediction of a completely known parameter [7]. On the contrary, in a real integration of data mining and CBM approach, prognostics is an integral part of the flowchart and makes use of the relevant features extracted in data mining step. This approach enables the possibility to compare different features evolution and combine them to improve the accuracy of remaining useful life forecast.

This approach consists of:

- A data mining step on historical data where data preparation, data reduction and relevant features extraction are performed.
- A prognostics step, where optimal thresholds are retrieved using prediction algorithms are applied on live data to estimate the remaining time for the relevant features to reach the thresholds. This remaining life time can be used to establish an optimal predictive maintenance.

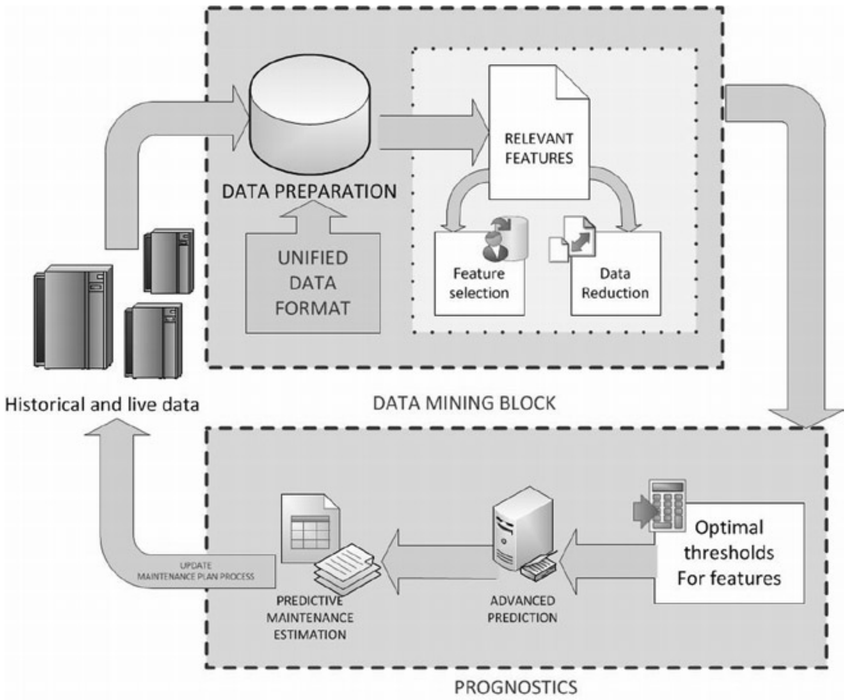


Figure 5. Data mining steps with maintenance data

Logical data mining process to optimally forecast the predictive maintenance actions is as follows: starting point is the available historical database of machines running whose maintenance performed actions can be altered in function of the result of decision support system regarding scheduling process. In general, such a database exists for almost all machine manufacturers, but only a limited amount of information is currently used.

The next step consists of data preparation, including data transformation from the original to unified data formats and data cleaning such as removing outliers or calculating missing values. Note that the data preparation step may be time consuming since the unified data format should be compatible with the data mining software tool and still retain a physical interpretation of the data.

The real data mining step consists of the data reduction where significant information is found and non relevant information is discarded to reduce significantly the number of attributes in the searching process. After this pruning process, the relevant features are extracted out of the selected.

The final step consists of prognostics, itself divided into two sub steps. First, reliability estimation methods are applied to historical database in order to identify the optimal thresholds of the selected features. Secondly, a prediction algorithm is applied to live data in order to forecast the time to schedule the optimal CBM. The different steps of the data mining approach are summarized in Figure 5.

5. Conclusion

Main benefit from maintenance data integration process is insight into the future establishment of CM and CMMS data format standards. Early integration attempts will identify data structures that are most conducive and useful for long-term storage and searching. With the coordination of CMMS and CM developers, a general set of guidelines for file formats can be established which will enhance the potential for research by the scientific community, greatly increasing the usefulness of CMMS and CM platforms. Following the implementation of an integrated system, the usefulness of CM devices will expand from mere guidance tools to automated diagnostics and prognostics systems. These integrated systems will constantly compare sensor readings to the wealth of historical records and forecast likely maintenance events based upon historical precedence. This will allow for more efficient logistics and component performance evaluation. Furthermore, these systems will identify unexplained or common modes of failure directing the efforts of scientific component testing, the results of which will drive design modifications making the assets increasingly more reliable.

This integration has two steps. First one is integration of technology and standards like MIMOSA are actively contributing to the development of a common hardware and software platform for data storage. Second step is related to the new knowledge extraction as a result of integration performed in first step. For this purpose data mining comes up as a very effective maintenance knowledge tool.

Data mining has become useful over the past decade in maintenance to gain more information, to have a better understanding of the behaviour of running assets, and to find optimal maintenance policies derived from this new knowledge. Today, data mining is no longer thought of as a set of stand-alone techniques, far from the maintenance applications. Enterprises require more and more integration of data mining technology with relational CMMS and CM databases and their business-oriented applications. To support this move to applications, data mining products are shifting from stand-alone, technology to being integrated in the relational databases.

The goal is to find new and interesting patterns in the data and somehow use the gained knowledge in maintenance decisions. The integration of mining results into the operational business is usually done in an ad-hoc manner. With the integration of mining, the focus shifts toward the deployment of mining in business applications. The audience includes the end users in the line of business who need an easy-to-use interface to the mining results and a tight integration with the existing environment.

Simple predictive models, anticipation of assets behaviour, and automatic optimization of maintenance processes by means of data mining become more important than some general knowledge discovery. If the purpose of the model is to increase knowledge of the data, the knowledge gained needs to be organized and presented in a way that the end user, maintenance personnel can use it.

Depending on the maintenance requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. However, even if the data analyst does not carry out the deployment effort, the maintenance personnel, i.e. end user, must understand up front what actions need to be carried out to actually use the created models. A right deployment of the model and understanding by maintenance people are necessary to be sure the real benefits are harvested by the company through right maintenance decisions. The final product should manifest itself as an automated maintenance exploration interface. Users should be able to quickly identify possible diagnoses of faults and quickly retrieve historical maintenance actions that were effective in resolving the problem. Such a system would be easily scalable allowing for maintainers to have information on a variety of practices being performed across the field.

In addition to the assurance that these systems deliver top performance, benefits of this integration include increased plant availability and lower maintenance costs because most faults are caught before they can evolve into problems requiring major repairs and/or costly process interruptions and downtime.

Acknowledgement: The author B. Tormos wish to thank "Programa de Apoyo a la Investigación y Desarrollo (PAID-00-11) de la Universitat Politècnica de València" for supporting his research.

References

1. Blair J, Shirkhodaie A. Diagnosis and prognosis of bearings using data mining and numerical visualization techniques. Proceedings of the 33rd Southeastern Symposium on System theory, 2001: 395–399
2. BRITISH STANDARD, BS 4778 (1991): Glossary of terms used in quality assurance (including reliability and maintainability). British Standards Institution, London
3. IEC (1990). 60050 (191): Dependability and quality of service. International Electrotechnical Commission, Geneva, Switzerland.
4. EN 15341: 2007, Maintenance Key Performance Indicators. European Committee for Standardization (CEN), Brussels, Belgium.
5. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS). UMETRICS, Sweden, 1999.
6. Frawley W, Piatetsky-Shapiro G, Matheus C. Knowledge discovery in databases: An overview. AI Magazine 1992; 13: 57–70
7. Goh K M, Tjahjono T, Subramaniam S. A review of research in manufacturing prognostics. 2006 IEEE International Conference on Industrial Informatics. pp.417–422.
8. Isermann R. Fault-diagnosis systems. Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors and Fault-tolerant Systems. New York, USA: Springer, 2006.
9. ISO 14224 (2006) Petroleum and natural gas industries – Collection and exchange of reliability and maintenance data for equipment. International Organization for Standardization, Geneva, Switzerland.

10. Jardine AKS, Lin D, Banjevic D A. Review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing 2006; 20:1483–510.
11. Kelly A. Maintenance Systems and Documentation. New York, USA: Butterworth-Heinemann, 2006.
12. Mather D. CMMS: a time saving implementation process. New York. USA: CRC Press, 2002.
13. Soibelman L, Kim H. Data preparation process for construction knowledge generation through knowledge discovery in databases. Journal of Computing in Civil Engineering 2002; 16 (1): 39–48.
14. Tianhao W, Khan F M, Fisher T A, Shuler L A, Pottenger W M. Posting Act Tagging Using Transformation-Based Learning. ICDM'02 1992 Proceedings of the Workshop on Foundations of Data Mining and Discovery, IEEE International Conference on Data Mining.
15. Wylie R, Mottola L, Kresta J, Monk R. Lessons Learned for the I02 Project. COM 2002 The Conference of Metalurgists (The International Symposium on Knowledge Management in the Metals Industry). Montréal, Québec, Canada.
16. Yang C, Létourneau S. Learning to Predict Train Wheel Failures. KDD 2005 Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, Illinois, USA.

Prof. Diego GALAR
Anna GUSTAFSON, MSc.
Division of Operation and Maintenance Engineering
Luleå University of Technology
Luleå, 97187 Sweden
E-mail: diego.galar@ltu.se
anna.gustafson@ltu.se

Prof. Bernardo TORMOS
CMT- Motores Termicos
Uiversitat Politècnica de València
Valencia, Spain
E-mail: betormos@mot.upv.es

Prof. Luis BERGES
Department Design Engineering and Manufacturing
University of Zaragoza
Zaragoza, Spain
E-mail: bergesl@unizar.es