



HAL
open science

CMBase, a universal gateway to condition monitoring datasets

Cécile Capdessus, Frédéric Bonnardot

► **To cite this version:**

Cécile Capdessus, Frédéric Bonnardot. CMBase, a universal gateway to condition monitoring datasets. Surveillance, Vishno and AVE conferences, INSA-Lyon, Université de Lyon, Jul 2019, Lyon, France. hal-02188545

HAL Id: hal-02188545

<https://hal.science/hal-02188545>

Submitted on 18 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CM Base, a universal gateway to condition monitoring datasets

Cécile Capdessus¹, Frédéric Bonnardot²

¹PRISME, Univ. Orléans, 21 Rue Loigny-la-Bataille 28000 Chartres, France
Cecile.capdessus@univ-orleans.fr

²Univ Lyon, UJM St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France
frederic.bonnardot@univ-st-etienne.fr

Abstract

Condition monitoring is a lively research domain, with hundreds of researchers sharing new techniques coping with complex problems, such as variability of the operating conditions, complex machinery monitoring, diagnostics, prognosis... Considering the many methods that have been proposed these last decades, it is striking how few of them have reached the industrial application. One of the reasons is that the intensive validation of these methods over real-life data required to ensure their reliability is often difficult to achieve. The results obtained over one dataset may not be easily reproduced over another one. Furthermore, can the comparison between two techniques be objective, when applied to different datasets?

Some laboratories and companies generously share the datasets recorded on their test benches or industrial devices. Unhappily, it is often difficult for them to know for certain who has worked on their data, which is both frustrating and a real problem when they report to their authority about the diffusion of the data, in order to get the subsidies required for the test bench exploitation. From the user's point of view, there can also be difficulties in finding the right datasets for a specific study among the jungle of all that is proposed on the web. How to find the appropriate dataset for a prognosis study, or a fatigue study, or the study of any specific kind of damage? Another problem can be met by those who wish to share their datasets with the scientific community but do not have the technical skills or staff to do so.

CM Base is a web portal that aims at facilitating the sharing of the data, by offering many a functionality, such as a list of all the existing datasets and test benches with all related papers, searching facility allowing to extract from the base all datasets related to a specific problem or papers related to a specific dataset.

1 Introduction

These last decades have seen a wide expansion of on line condition monitoring, due to various factors. First, the evolution of sensors, computing and communication technologies allow setting up condition monitoring systems more easily and at lower cost. Second, the need of such systems has increased. Indeed, in the present competitive international context, production companies cannot bear the consequences of unplanned production lines shutdown. Furthermore in many domains, products are lent rather than bought by the clients, or bought with some service package, including monitoring. The development of machine learning and Artificial Intelligence (AI) has also opened new fields for automated surveillance.

In this context, the need for reliable monitoring techniques has led industry as well as academic researchers to explore many possibilities, often based on the use of a set of sensors and the elaboration of relevant health indicators through the application of signal processing techniques. The elaboration and the testing of these methods require the production of experimental data.

Many laboratories have set test benches and recorded measure signals datasets and some of them chose to share these datasets with the scientific community through a web site. But it reveals difficult for them to get information about the diffusion of their dataset: for instance who it has been used by, the conferences and papers the obtained results have been published in. For their part, the users have some difficulties in finding the right dataset to test their technique or even to know about the available ones. Furthermore, some laboratories wish to share their datasets but lack computer knowledge or facility to do so.

CM Base is a database meant both to gather all the information about existing on line datasets and to help laboratories putting online their datasets. It can be accessed through a web site that allows inserting a new

dataset, inserting a paper related to an existing dataset, and browsing the basis in order to find some datasets relevant to a specific problematic with all related papers. In order for the basis to be as widely used as possible, the interface has been designed to simplify all procedures, which makes CM Base a very easy to use application.

In section 2, the need is analysed and a solution is designed, in section 3 the chosen technical solution is presented, in section 4 a user guide for the use of CM Base is given, and in section 5 a conclusion and some perspective are proposed.

2 Analysis of need

2.1 Reliability and reproducibility

Reliability and reproducibility of the techniques have become one of the main topics in all scientific domains [1], peculiarly in all that is related to data processing and statistical analysis. Indeed, for a specific technique to be validated, it is not enough that it has been tested on one set of data. The quality of the obtained results can depend of some parameters related to the specific experimental setup, the location of the sensors, all things that are different from a test bench to another. Furthermore, real-life signals are yet another challenge. Techniques that have been tested on measure signals collected from a lab test bench can lack robustness when it comes to measures collected from an industrial device [2, 3]. It is thus of utmost importance that researchers might access to a wide set of different measures, in order to test the robustness of their proposed techniques and to specify in which conditions these can be applied. Researchers should thus be given access to a wide range of measure signals, collected on both lab test benches and real-life devices.

2.2 Comparison of techniques and measure signals

No indicator can be considered perfect for the monitoring of a system. Depending of its stage, the presence of some damage may produce on the measure signal alterations of a different nature. For instance, a localised defect on a roller bearing can be detected with techniques well adapted to transient signals, whereas once the damage has spread, such techniques will fail to put it at light. Furthermore, different kind of measure signals can be used to monitor a system: current, vibrations, temperature... For these reasons, most condition monitoring AI systems use both different kinds of measure signals and different kind of indicators. It is thus interesting to compare either the results obtained through one specific method on different datasets, or through different methods applied to a specific dataset. This can help defining how to combine these techniques and merge different indicators in order to ensure a reliable monitoring system. It would thus be interesting for the condition monitoring community to have access not only to different kind of measure signals, but also to the related bibliography.

2.3 Choice of an appropriate dataset

When addressing some specific topic that may be bearings, gearboxes, diagnosis, prognosis, wear, variable load, variable speed or any other relevant topic for condition monitoring, accessing to a measure dataset well adapted to this topic is not an easy task. All labs are not equipped so as to set up an experiment and while many datasets are now available on the web, it takes time to check whether they have the right profile. It would be convenient that all these datasets can be accessed through a browsing system that would allow getting straight access to the well-fitted ones. And then, for each dataset, it would be interesting to have also access to the related state of the art. On the top of it, it would ensure to the researchers who have already published on this dataset that any newcomer is informed of their work.

2.4 Storage and feedback

Many labs and even companies have chosen to share their datasets through their lab webpages or at the occasion of a conference contest [4]. In order to justify the maintaining of their test benches, they must prove to their organisation that sharing the data with the scientific community is profitable. This profit can be expressed in terms of international influence or appropriation of new techniques. The problem is that they often get no more than the number of times their data have been downloaded, or at best who asked for them. They rarely get information about what has really been done on their data. They sometimes discover by chance that their team is thanked in a paper for having shared the data. It would be interesting for them to get a list of all papers that have been published about techniques that were applied to their dataset. In other words, it would be useful to try to each dataset the related bibliography. Furthermore, some researchers would like to share their data with the community but have no technical means to do so. It would be interesting to provide the community with a storage facility to share their dataset.

2.5 Specifications

The application should allow:

- Sharing a dataset ;
- Sharing the reference of a publication related to a dataset ;
- Browsing the base in order to get a list of all datasets relevant to:
 - A specific topic (diagnosis, prognosis...);
 - A specific device (gearbox, bearing, engine...);
 - A specific context (lab test bench, industrial recordings);
 - Specific operating conditions (constant or variable load or speed for instance);
- Getting the bibliography related to any dataset.

The access should be authorized only to registered visitors, so that the application must include some access management part. The interface must be easy to use, so that contributors do not lose time whenever they add a dataset or a paper, or browse the base.

3 Technical solution

3.1 Strategy of development

CM Base is accessed through a web site. This web site can be used in two ways: as a data navigator to search data by keywords and obtain associated bibliography, or as an editor to declare existing data and databases, associated articles and documents.

The catalogue of data, database, associated links, keywords and bibliography are stored in a relational database. The associated documents and pictures are directly stored on the server. CM Base is not a data storage server (like dropbox) that manages hundreds of MegaBytes of data, but a gate allowing access to the data. Only links to data and small document are stored into the database. The user can use a file-sharing solution such as Mendeley [5] to store his measure datasets, and then declare these data in CM Base to give both acces to the data and information about it: keywords, little document, link to files, ...

The website was developed by using standard LAMP tools: Apache, MySql and Php and a Linux Debian web server hosted by Université Jean Monnet of St Etienne on a Virtual Machine.

In order to provide a long-term solution not sensitive to update we avoided using a web Content Management System (CMS) like Wordpress, Drupal, Joomla, ... The solution was directly coded in php [6]. In the same way, instead of using css toolkits like bootstrap that require extra files and regular update, a more classical css formatting was used , based on flexbox.

A Model-View-Controller architecture [7,8] was used to organise the project. The database was designed through a Merise analysis [9] and is managed by sql requests in Php.

3.2 Database structure

The organisation of the database is illustrated by figure 1. Each box corresponds to a table, which is equivalent to a sheet from a spreadsheet. The name at the top is the name of the table/entity. The following elements are the fields, which are the equivalent to the column names in a spreadsheet. The interest of relational database is that it allows establishing some links between the tables. These links are symbolised by rounded boxes. For example, all fields related to a member, i.e. his member_ID, last name, e-mail, and so on, are stored in a Members Table/Entity. A link is established with the Papers Table in order to associate to each member the publications he introduced into CM Base. This link is created by adding a (table_ID, member_ID) pair to the IntroducedBy Table. More technical details about this representation could be found by searching literature on Merise method for Database Analysis.

The main tables are:

- The Members table, to store a list of people registered in CM Base,
- The Papers table that contains a list of papers stored in the database,
- The Datasets table, to store information about each test bench or data collected in the database,
- The Ressources table, in which are collected a list of files, links, pictures.

The links between tables enable us to find articles associated with datasets (About relation) or to make a link between type of Sensors and Datasets. The use of a database facilitates data management since all creation, updates of data are made by requests to MySQL server, so that all the storage of the information is managed by Mysql.

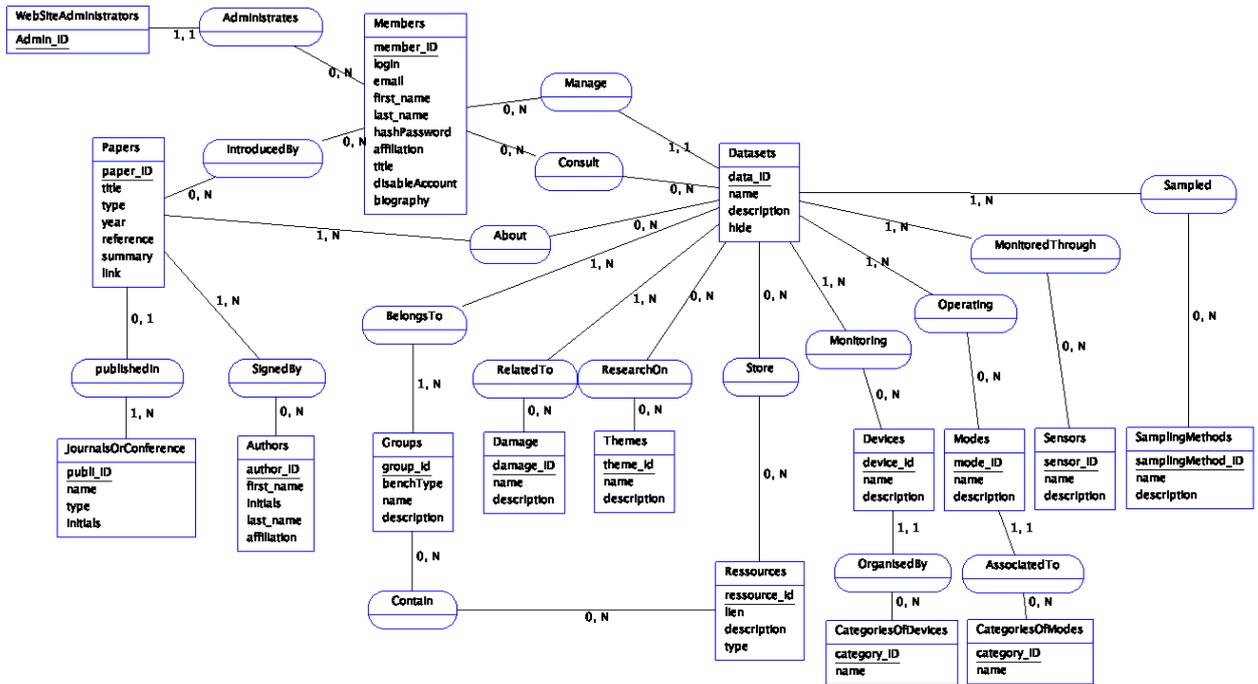


Figure 1: Database Conceptual Model

3.3 Organisation of the Web site

Access to the data is provided through a web site. A menu enables accessing to the following pages:

- “Search the base” to browse inside the collection of data,
- “Add my data” that enables the user to insert its own data into the database or made some links with other databases,
- “Add my papers” to declare papers linked with a dataset,
- “My profile” to change or give personal details,
- “My contributions” to list papers and data inserted by the logged user.

The next section will present some of the web pages as an example.

4 User guide

In this user guide are presented some of the functionality of CM Base. A more exhaustive user guide is available on CM Base web site.

4.1 First connection

Figure 2 shows the home of the database. An account must be created for the first connection by using request account link in the menu. The request account page asks some personal details, an e-mail, a biography and a password.

After connection, more options are available on the menu as shown in figure 3.

4.2 Search the base

The “search the base” web page lists the content of the dataset and enables to make search by using tags associated with the dataset (sensors, device, type of sampling, ...). These tags are the same as those used for the dataset insertion (see “Add my data” section).



Figure 2: Home page

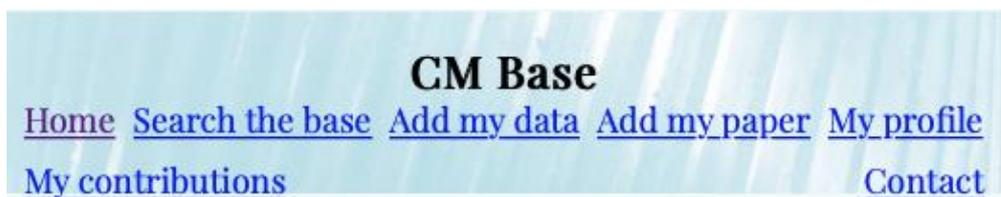


Figure 3: CM Base menu

4.3 Add my data

The menu item “add my data” allows inserting your data into CM Base and sharing them. The associated page is shown in Figure 4. At first, only the “Description of the dataset” part is displayed. Once this part is completed and validated by clicking on the “Create Dataset” button, the next elements appear.

The description of the dataset and main information sections display the same elements as those displayed in a dataset list or dataset search answer. Each element can be modified after submission if necessary.

The user should provide a dataset name, a short description (short enough to be displayed in a table of answer). A longer description can be provided as a pdf file in the main information file.

The six thematic lists are very important and should be filled with care, since the selected items will be used to perform sharper browsing through the base. One or more elements can be selected.

The “Devices” theme enables identifying the component of interest (gear, rolling elements bearing, ...) within the dataset and the associated machinery. It is also necessary to give information about the damage, the sensors and sampling methods. The “Mode” section allows giving extra information on running conditions: speed, load, ... Three different acquisition modes are available: diagnostic when it is necessary to find the type of defect, prognostic when it is necessary to estimate a lifetime duration, and, data mining when a large set of data with different states (health, damage, ...) are provided.

The “Access to data” section gives three alternatives to provide data (it is possible to combine these options): first, the data can be obtained through an e-mail request to the associated contact, second, one or more links to a file on a server can be provided, and third, if a web page associated with the data already exists, it is possible to provide some web links.

Additional files like calibration information, document of the test bench, scripts to read the data can also be provided.

Propose Data

Description of the Dataset

Give a name, a short description of your dataset, some information about its contents, and press Create dataset.

Name of the dataset :

Short description :

You can select multiples elements in the list if you hold down the Ctrl (or cmd on a mac) button.

Devices : Other Other Component of interest Gears Rolling Elements Bearing Belt Blade Rotor / Stator Whole machinery Machining centre Wind turbine Diesel Engine Electrical Engine Pump Context Industrial Laboratory	Problems : Other Pitting Crack Corrosion Imbalance Misalignment Cavitation	Modes : Other Other Speed Constant speed Linear speed Non stationary speed Very Low speed Very High speed Load Unload Nominal load	Sensors : Other Accelerometer Voltage Current Torque-meter Force Optical encoder Microphone Temperature	Sampling Method : Other Temporal Sampling Angular sampling Non Uniform Sampling Elapsed Time	Theme : Other Data Mining Diagnostic Prognostic Run to Fail
--	--	--	---	--	---

Create dataset

Ressources : Main information

Provide a logo and main description / documentation pdf file - press Upload after each file selection

These files will appear in search results. More pictures and files can be provided in the "Various files" section.

Detailed description : aucun fichier sél. (only one pdf file - maximum size 10 MB)

Logo (small picture) : aucun fichier sél. (only one jpg or png file - maximum size 1 MB)

Ressources : Access to data

This website does not directly store data but links to data.

You have many options to share your data : (choose one or more)

- I require that each user contact me be e-mail :
- I give link(s) to my data (stored on a server) :
 - web address : (do not forget ftp:// or ssh:// or ..)
 - description :
 -
- I give link(s) to my website(s) :
 - web address : (do not forget http:// or https://)
 - description :
 -

Ressources : Various files

You can upload here a collection of additional documents or pictures.

Add documents :

- aucun fichier sél. (Maximum size 10 MB)

Add a picture :

- aucun fichier sél. (maximum size 5 MB)

Figure 4: Add my data

4.4 Add my paper

Every user can declare paper associated to a dataset by using the "Add my paper" page. To do so, simply paste the citation of your paper in a format similar to bibliography style and click next. CM Base will analyse your entry and try to identify the author, the title, the journal and so on. The result of this analysis will be presented in a web page and you will be proposed to correct the values and to link your paper to a dataset.

4.5 My Profile

My profile web page allows entering or modifying your personal details, biography, password or delete your account.

4.6 My Contributions

This web page shows a list of submitted datasets as well as associated bibliography. It also gives the possibility to edit or delete the dataset. Since this database is moderated, the dataset should be accepted by the webmaster before being viewable on the web site.

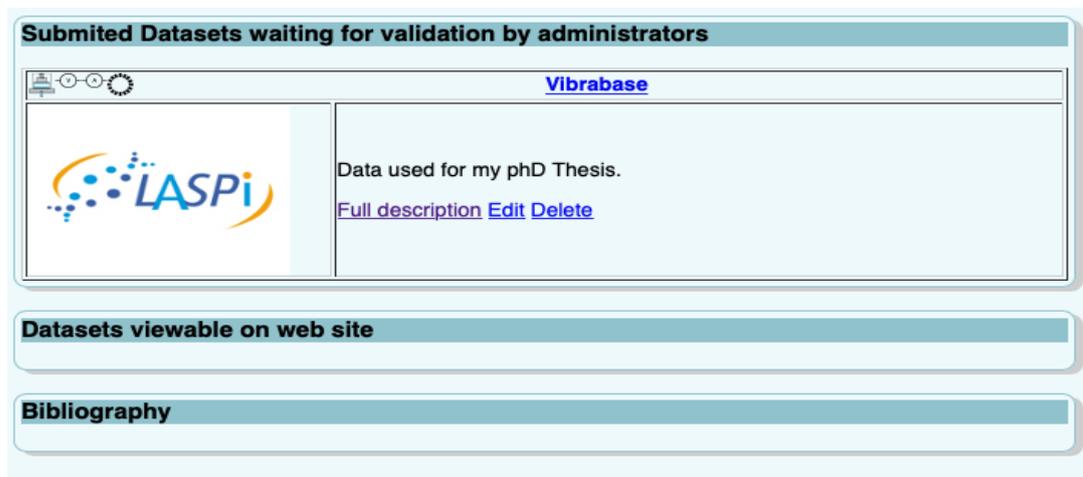


Figure 5: Add my data

5 Conclusion

By proposing a Condition Monitoring data base, CM Base, we aim at filling some gap in our research community. The proposed application will allow finding datasets well-fitted to a specific study, related publications, proposing new datasets and sharing publications. It was designed to be easy to use, and easy to maintain. The structure is such that some new functionality can be added if requested by the users. Some suggestions have already been made: auto alert messages, forums to debate about datasets and methods, challenges proposed by companies on hot topics, groups of datasets used for contest or teaching purpose, overview publications and so on. The interest of the application will depend on its success. The more datasets and papers will be referred to in the base the wider its use will spread. We thus count on your contributions and suggestions to improve the application.

References

- [1] M. Baker, Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help, *Nature*, vol. 533, no. 7604, 2016, p. 452+.
- [2] T. Barszcz, How to bridge academia and industry ? A practitioner's perspective, Keynote lecture, *Surveillance 8*, Roanne, France, October 20-21, 2015.
- [3] R. Klein, *Automatic diagnostics of complex machinery*, Keynote lecture, *Surveillance 8*, Roanne, France, October 20-21, 2015.
- [4] J. Antoni et al, Feedback on the *Surveillance 8* challenge: Vibration-based diagnosis of a *Safran aircraft engine*, *MSSP*, Volume 97, December 2017, Pages 112-144.
- [5] <https://data.mendeley.com/>
- [6] PHP Notes for Professionals, Free Programming Books, goalKicker.com
- [7] G. E. Krasner, S.T. Pope, A Cookbook for Using the Model-View-Controller User Interface Paradigm in Smalltalk-80, *JOOP*, September 1988.
- [8] T. Reenskauag, *The original MVC reports*, thing-model-view-editor, 1979.
- [9] J. Gabay, *Merise et ULM pour la modélisation des systèmes d'information*, Dunod, 5^{ème} édition, 2004, ISBN 2 10 007205 6.