2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

National Metrology Laboratory of the Philippines Digital Calibration Certificate: DigiCert

Ahdrian Camilo C. Gernale[1], Nathaniel Ken A. Aquino[2], Roj Gian D. Gorospe[3], Mark Joseph C. Nicasio[4]

[1]*Industrial Technology Development Institute, Taguig City, accgernale@itdi.dost.gov.ph*
[1]*Asian Institute of Management, agernale.PhDinDS2028@aim.edu*

[2] *University of Santo Tomas, Manila City, nka.aquino02@gmail.com*

[3] *University of Santo Tomas, Manila City, gorosperojgian@gmail.com*

[4] *University of Santo Tomas, Manila City, nicasio.markjoseph641@gmail.com*

Abstract— This paper presents the design and implementation of a system application named *DigiCert*, a Digital Calibration Certificate (DCC) application developed by the National Metrology Laboratory of the Philippines (NMLPhil) of the Industrial Technology Development Institute to modernize calibration workflows and enhance certificate management through standardized, machine-readable documentation. DigiCert improves the process of changing old formats like scanned images and PDF files into organized XML using the Physikalisch-Technische Bundesanstalt (PTB) DCC XML schema. It also allows for manual data entry with a real-time XML preview and can convert back and forth between XML and an easy-to-read PDF. The system employs advanced techniques to enhance images, detect text arranged in a grid pattern, and recognize characters in order to extract both tabular and free-form data from those images. For PDF, it uses Python libraries like pdfplumber and regex parsing to ensure high accuracy. A user-friendly graphical user interface guides the operator through import, edit, and export functions across five dedicated windows, enforcing DCC schema compliance and reducing administrative overhead. DigiCert's flexible design and compliance with ISO/IEC 17025:2018 standards guarantee that the application can work well with other systems, keep track of changes, and maintain security through cryptographic signatures, making it a suitable choice for modern digital measurement and certified calibration labs.

I. INTRODUCTION

A. *What is a Digital Calibration Certificate?*

A DCC is an electronic version of a traditional calibration certificate that provides standardized, machine-readable documentation of a device's calibration. It is part of the digital transformation in metrology (the science of measurement), aiming to improve accuracy, traceability, interoperability, and efficiency in calibration processes [1].

The DCC uses the widely accepted XML format, making it machine-readable and easily integrated into digital processes. It uses trusted digital signatures, which are created and checked with established security methods, to ensure that a DCC is genuine, has not been altered, and cannot be denied by the sender. Various government systems already rely on these security methods. Additionally, the DCC complies with the DIN EN ISO/IEC 17025:2018-03 standard, making it suitable for use in accredited laboratories [1].

Currently, the PTB, Germany's National Metrology Institute, has developed a web application called the GEMIMEG-Tool, which allows users to generate DCCs by inputting the required data. This tool is publicly accessible through PTB's website.

The goal of this project is for NMLPhil to develop its own DCC application and integrate it into the organization's routine operations. The application, named **DigiCert**, aims to streamline and accelerate the calibration process while aligning with advancements in digital technology and improving certificate management by enabling digital storage.

DigiCert comes equipped with several key features:

1. Convert existing file formats (e.g., .jpg, .pdf) into XML to minimize manual data entry.
2. Supports manual input and editing of certificate data for flexibility.
3. The system has the ability to transform XML files into a format that is easily readable by humans or PDF, thereby guaranteeing accessibility and ease of use.

The design of these capabilities aims to offer a more efficient, user-friendly, and future-ready solution for managing calibration certificates.

B. Benefits of DCC

DCC allows for calibration certificates to be **machine-readable**, allowing for seamless integration with various systems that can operate on the data gathered from the calibrations [3]. The DCC follows a tree-like structure, organizing the data hierarchically, which makes it easier for machines to piece together information.

It also enhances the **traceability** and ensures better tracking of the measurements and standards used during the calibration. This digital traceability not only supports compliance with the standards to meet quality requirements but also increases the validity and reliability of the results gathered in the calibration process [2].

An XML format for the calibration certificates also increases the efficiency of gathering data and speeds up the process of uploading, validating, and using the data. Additionally, it reduces the administrative overhead and minimizes the amount of paperwork required for processing.

An XML format also increases the interoperability and compatibility of calibration certificates. Its structure enables the data to be placed in a hierarchical structure, which allows for better data placement and could be used for data sharing across laboratories, customers, and regulatory agencies [2].

C. Samples of Valid Certificates

One of the current limitations of the application is that it is primarily optimized to read and recognize calibration certificates issued by NMLPhil, particularly certificates from the Force Laboratory. As a result, its compatibility with certificates from other laboratories or formats may be limited at this stage.

D. Scope and Limitations of the Application

Given that the application is in its initial development phase, it comes with certain defined scopes and inherent limitations.

1. The application is currently optimized to recognize NMD's Force Calibration Certificates. Other NMD laboratories can still use it, but it may not fully support all inputs at this stage.
2. PDF inputs will have better results in reading data than image inputs because of OCR limitations.
3. The XML-to-PDF feature of the application currently does not support XML calibration certificates from other laboratories, such as international institutions, as of this moment. It can only process XML files generated through the application's own XML export feature.
4. The XML-to-PDF conversion is not a one-to-one replica of the original certificate; its primary purpose is simply to make the content human-readable.

II. DCC Schema

The DCC follows a schema developed by PTB, which is available on their website. The development of the application primarily used this schema as a guideline, although it served as a reference. Calibration certificates don't have a strict format; they tend to be similar or closely aligned. The schema used in the application was based on PTB's standard but was adapted and structured to better align with NMLPhil's calibration certificate format.

The project DigiCert was developed in compliance with this schema to ensure standardization and compatibility with digital calibration frameworks.

At the top level is the digitalCalibrationCertificate element, which is of type dcc:digitalCalibrationCertificateType. This serves as the root element, containing all the main sections (also referred to as the "four rings" of the DCC) [1]:

1. Administrative Data (administrativeData) – Contains important metadata about the calibration process such as customer information, certificate number, and laboratory details.
2. Measurement Results (measurementResults) – Holds the measurement data, including values recorded during the calibration and their associated uncertainties.
3. Comment (comment) – An optional field that allows for human-readable notes or remarks.
4. Document (document) – Stores supporting documents or binary data relevant to the calibration, represented by the byteDataType.
5. Additionally, the schemaVersion attribute provides versioning for the DCC structure, allowing systems to interpret the document format correctly.

By following this structure, the DigiCert application ensures that every XML output it generates is aligned with international standards for digital calibration certificates.

III. Methodology

The entire DigiCert application was developed using the Python programming language, utilizing various libraries to ensure full functionality and seamless integration of its features. The following sections describe how the various features of the application were developed and implemented.

A. Image to XML

A primary function of the application for converting calibration certificates to an XML (DCC) format is to convert an image of a certificate to the DCC format. By passing an image as input to the application, the system can read the data from it, convert it to the structure of the XML schema, and allow the user to add or edit more details before exporting the calibration certificate as an XML file.

1. Approach used for Optical Character Recognition (OCR)

There are three main steps in reading the calibration data from images to turn them into texts for the system to be able to process and convert them to the XML format. The steps taken include image preprocessing, text detection, and text recognition. These steps were followed sequentially to ensure that most of the calibration details are able to be properly read and be ready for conversion to XML.

Preprocessing the input images comes first before going through the next steps. This utilizes several image-enhancing algorithms to improve the quality and clarity of the images before the main steps of optical character recognition are applied. The preprocessing of the image includes several changes, such as removing noise, combining color channels for better clarity, adjusting contrast with the Contrast Limited Adaptive Histogram Equalization algorithm (CLAHE), and using Gaussian blurring. CLAHE is an image processing technique that enhances the contrast of images, particularly in areas where contrast is low. Denoising reduces the color noise while preserving edges, which makes the texts more distinguishable from the background. Using Lab color space, also known as CIELAB, separates color information into lightness (L) and two color channels (a and b). The L channel represents the lightness of a color, ranging from 0 (black) to 100 (white). Additionally, the image is changed to the Lab color space, which separates lightness (L) from color (A and B). This process helps improve contrast, especially in the lightness channel, making faded or uneven text easier to see. Finally, Gaussian blur and sharpening are also used, wherein a blurred version of the image is subtracted from the original (using weighted difference) to enhance edges and other fine details in the document, resulting in clearer character strokes for the OCR to recognize.

The next step is the text detection, wherein the goal is to identify the regions of interest in the image, particularly tabular or grid-based layouts that are likely to contain text. To find these areas, the image is first changed to grayscale and then inverted using adaptive thresholding to help separate the text and lines (foreground) from the background. Following this the horizontal and vertical line structures are then extracted using erosion and dilation with shaped kernels to help in isolating table lines or cell boundaries. Using the detected lines, a grid-like mask is then constructed by combining the horizontal and vertical lines. Contours of the rectangular regions are then extracted. The detected cells are then filtered based on their respective area and calculated median threshold to remove noise and to avoid detecting the cell's table as another "cell," as it would duplicate the OCR on the table. The bounding boxes are sorted top-to-bottom and left-to-right to preserve the natural reading order.

Finally, for the OCR, we used a machine OCR algorithm called Tesseract. This employs a two-stage adaptive recognition process. It begins by identifying character patterns, then uses context to refine its understanding of words and sentences, and then converts image regions into actual text. We used several Tesseract configurations, such as *--oem 3*, *--psm 12*, and *--l eng*. The configuration *--oem 3* utilizes the LSTM-based OCR engine, which provides high accuracy for both printed and handwritten texts that are the target input of the DCC system. Meanwhile, *--psm 12* specifies the type of page segmentation, treating the image as sparse text with possible orientation and script detection. As the calibration certificates are a combination of descriptions, tables for results, and paragraphs, *--psm 12* was chosen as the best configuration. Finally, *--l eng* tells Tesseract that the expected language is set to only English. If the DCC system expands to other languages, this configuration can be deleted.

2. Using OCR on a scanned document

After running the OCR on the system, here are the generated results saved on a temporary PDF file to be converted into an XML format:

For the scanned images, the OCR system can better detect texts as there is less environmental noise that is happening in the image. Additionally, the foreground is already quite differentiable from the background.

Images captured with a camera could result in a less accurate OCR, as there are many outside factors that could add noise to the image. Even a blurred or low-quality image can affect the quality of the text recognition system. However, the image still retains most of its details.

B. PDF To XML

The conversion of calibration certificates from PDF to XML ensures accuracy and reliability in extracting data from text-based PDF documents since the application leverages the pdfplumber Python library.

The Python library pdfplumber is specifically designed for extracting text, tables, and metadata from text-based PDF documents. It is particularly effective for digital PDFs that contain selectable text, rather than scanned image-based documents. Using this Python library, the application is able to accurately retrieve paragraph-level and line-by-line text from PDF pages and seamlessly process PDFs with multiple pages.

The process begins when the user selects a PDF file through a file dialog interface. Once the file is chosen, it is opened using the pdfplumber.open() function, which loads all the pages of the PDF. The application then loops through each page, using the extract_text() method to retrieve the textual content. All the extracted text from the pages is combined into one string called raw_text, which serves as the foundation for further data extraction.

After obtaining the text, the application extracts tables from each page using the extract_tables() method. Each table is processed by assuming that the first row contains the headers, while the following rows represent the data entries. These tables are stored in a dictionary format where each header is used as a key, and the corresponding values are the rows of data beneath it. This structure allows for easier access and organization of tabular information.

Specific pieces of information needed for calibration records are extracted from the raw text using regular expressions. This includes details such as the certificate number, calibration date, customer name, and address. For example, a pattern like CERT-\d+ is used to find the certificate number. The same technique is used to extract other important fields, and additional logic is applied when the formatting of the certificate varies across files.

Once all relevant text and table data are extracted, they are organized into a structured Python dictionary called calibration_info. This dictionary uses clear and consistent key names such as "certificate_number," "calibration_date," and "customer_name," with each key assigned its corresponding value. This structure prepares the data for easy conversion into other formats.

The dictionary is then saved as a JSON file using Python's built-in json.dump() function. Saving the data as JSON ensures that it remains structured, readable, and ready for further processing. This step also helps separate the data extraction phase from the XML conversion phase, allowing greater modularity in the system.

Finally, the JSON file is sent to a separate script responsible for generating the XML output. This script reads the JSON data using json.load() and constructs an XML structure. Each field in the JSON dictionary is turned into a corresponding XML element. The resulting XML file is then saved and formatted according to the required schema, making it suitable for submission, integration, or storage.

C. Creating New DCC

The creation of New DCC is designed to simplify the manual creation of XML files for calibration records through an interactive graphical interface. The process begins with the initialization of the application, where static configuration details, such as laboratory information and software metadata, are loaded from a JSON file.

When the application starts, it builds a simple and user-friendly interface. On the left, users enter details like certificate number, date, equipment, and customer info using text boxes and drop-downs. On the right, a live preview shows the XML output as data is entered, helping users check their inputs right away.

As users enter values into the input fields, the script collects all relevant data and compiles it into a structured dictionary. This dictionary is then used in the XML generation phase, where a predefined XML template is loaded and populated with the input data. Proper XML namespaces are handled to maintain schema compatibility.

The real-time preview updates dynamically as users edit the form, providing an accurate representation of what the final XML will look like. Once all information is complete, the user can export the XML file by choosing a save location through a file dialog. A confirmation message appears after a successful export. Users are then navigated back to the main menu, allowing them to start a new entry or perform other tasks.

This structured and intuitive process ensures that users can generate consistent, accurate XML files for calibration data without needing to manually code the XML structure themselves.

D. Conversion process of XML to PDF

This is the overall process of converting the XML-generated DCC of DigiCert back to PDF. The process starts by setting up the script, which includes defining XML namespaces and creating a helper function to safely extract text from XML elements. The XML file is then loaded and parsed. Using XPath queries, important data is pulled from specific parts of the XML, such as software details, administrative information, equipment used, lab details, responsible individuals, customer data, and measurement results. XPath stands for XML Path Language, which is a non-XML syntax to provide a flexible way of addressing (pointing to) different parts of an XML document [4].

Next, the content for the PDF is organized into sections. Each section is added to a list called the "story," which holds everything that will go into the PDF. A special table is created for the measurement results to make the data easier to read.

Once all the content is ready, the PDF is generated using ReportLab's SimpleDocTemplate. Finally, the finished PDF file is saved to the user's chosen location. This ensures that all XML data is clearly and correctly converted into a well-structured PDF document.

IV. The Application Result

Since the output of the project was an application, its graphical user interface (GUI) was designed to be interactive and user-friendly to ensure optimal usability, as shown in Figure 1. The application consists of five main functional windows. This section presents and describes each window in detail.

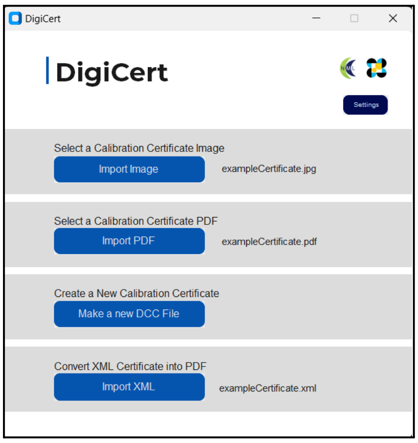


Figure 1. MAIN WINDOW

The main window greets the user upon launching the DigiCert application. The main window functions as the central hub, providing access to all the application's core features and functionalities. As previously mentioned, the app allows users to scan and convert image files to XML, convert PDF files to XML, create a new XML file from scratch or manually input data, and finally, convert an XML certificate into a human-readable format (PDF). All of these features will be discussed in further detail in the succeeding sections of this paper.

The "Settings" button is located on the upper-right portion of the main window. When clicked, it redirects the user to the settings window, where they can edit or modify certain information within the application.

This is the layout of the settings window. It contains static values such as country, language, and laboratory information. These static values are typically constant and do not require frequent changes. However, users still have the option to modify them when necessary. By default, the fields are pre-filled with standard values as shown above. After making any adjustments, users can click the “Save” button. The changes will then be applied and reflected accordingly upon returning to the main window.

Each text box provides visual feedback: it will appear red if a required field is empty and green once valid input is entered. This helps guide users to complete all necessary fields. The form is scrollable, allowing users to review and verify the entire certificate content.

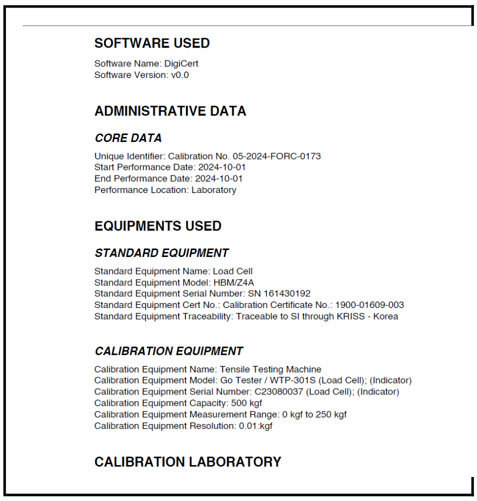
It is strongly recommended that users follow the format indicated by the placeholder text in each field. This ensures that the resulting XML file adheres to the DCC standard and maintains consistency.

Once all information has been reviewed and finalized, users can click the "Export" button to generate the XML version of the certificate in Figure 2.

The New XML window shares the same layout and appearance as the other windows in the application. However, the key difference is that users must manually input all the required information from scratch. Unlike the image or PDF import features, no data is automatically extracted in this window, so users are expected to fill in each field completely.

Figure 2. EXPORT

Figure 3 shows what the export notification results are when the XML file is exported. Once the export is complete, the user will be redirected back to the main window.



SOFTWARE USED

Software Name: DigiCert
Software Version: v0.0

ADMINISTRATIVE DATA

CORE DATA

Unique Identifier: Calibration No. 05-2024-FORC-0173
Start Performance Date: 2024-10-01
End Performance Date: 2024-10-01
Performance Location: Laboratory

EQUIPMENTS USED

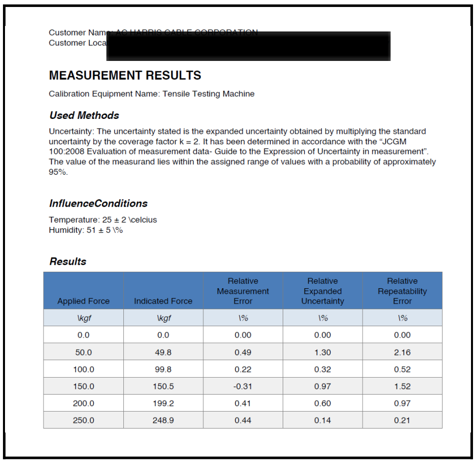
STANDARD EQUIPMENT

Standard Equipment Name: Load Cell
Standard Equipment Model: HBM/Z4A
Standard Equipment Serial Number: SN 161430192
Standard Equipment Cert No.: Calibration Certificate No.: 1900-01609-003
Standard Equipment Traceability: Traceable to SI through KRISS - Korea

CALIBRATION EQUIPMENT

Calibration Equipment Name: Tensile Testing Machine
Calibration Equipment Model: Go Tester / WTP-301S (Load Cell); (Indicator)
Calibration Equipment Serial Number: C23080037 (Load Cell); (Indicator)
Calibration Equipment Capacity: 500 kgf
Calibration Equipment Measurement Range: 0 kgf to 250 kgf
Calibration Equipment Resolution: 0.01 kgf

CALIBRATION LABORATORY



Customer Nam[redacted]
Customer Loca[redacted]

MEASUREMENT RESULTS

Calibration Equipment Name: Tensile Testing Machine

Used Methods

Uncertainty: The uncertainty stated is the expanded uncertainty obtained by multiplying the standard uncertainty by the coverage factor k = 2. It has been determined in accordance with the "JCGM 100:2008 Evaluation of measurement data- Guide to the Expression of Uncertainty in measurement". The value of the measurand lies within the assigned range of values with a probability of approximately 95%.

InfluenceConditions

Temperature: 25 ± 2 \celcius
Humidity: 51 ± 5 \%

Results

Applied Force	Indicated Force	Relative Measurement Error	Relative Expanded Uncertainty	Relative Repeatability Error
\kgf	\kgf	\%	\%	\%
0.0	0.0	0.00	0.00	0.00
50.0	49.8	0.49	1.30	2.16
100.0	99.8	0.22	0.32	0.52
150.0	150.5	-0.31	0.97	1.52
200.0	199.2	0.41	0.60	0.97
250.0	248.9	0.44	0.14	0.21

Figure 3. SAMPLE PDF FILE

Finally, users can also use the XML to PDF feature to transform the generated XML into a human-readable format. A PDF file will be exported containing information in the XML.

V. Conclusion

To improve certificate management and modernize the reporting of calibration through standardized, machine-readable documentation, NMLPhil created DigiCert, a powerful DCC application, which was presented in this paper. The primary goals and accomplishments of the system include

1. DigiCert efficiently operationalizes the conversion of conventional scanned images and PDF report files into PTB-compliant DCC XML.
2. Using smart OCR technology for images and accurate PDF reading with Python tools, DigiCert has demonstrated a dependable way to convert between XML and easy-to-read formats while ensuring both data accuracy and user-friendliness.
3. The application features a five-window GUI that guides users through every stage of import, editing, and export, and instant XML previews that enforce schema compliance

REFERENCES

[1] Digital Calibration Certificate - DCC - PTB.de. (2024, March 26). https://www.ptb.de/cms/en/research-development/ptbs-innovation-clusters/innovation-cluster-for-digitalization/kernziel1einheitlichkeitim/digital-calibration-certificate-dcc.html

[2] Koval, M., Tesař, J., Schönhals, S., Sharma, A., Balslev-Harder, D., Galvan-Hernandez, C., Baumann, H., & Mahovský, D. (2025). Digital Calibration Certificates (DCC): Benefits and Challenges to overcome for implementation [Conference paper]. EPJ Web of Conferences, 323, Article 01001. https://doi.org/10.1051/epjconf/202532301001

[3] Bernien, M., Straka, M., Bock, T., & Niepraschk, R. (2025). Digital calibration certificates and automated calibration workflow in the Vacuum Metrology Group of PTB. Measurement: Sensors, 38 (Supplement), 101490. https://doi.org/10.1016/j.measen.2024.101490

[4] https://developer.mozilla.org/en-US/docs/Web/XML/XPath