An Analysis of the Calibration and Measurement Capabilities from the National Metrology Institutes under the CIPM Mutual Recognition Arrangement; How Do the Laboratories Compare? Is This What Users Want?

Speaker: Diane Beauvais
International Office
National Physical Laboratory
Queens Road, Teddington, Middlesex TW11 0LW
United Kingdom
Phone: +44 20 8943 6477; Fax: +44 20 8943 6079
E-mail: diane.beauvais@npl.co.uk

Authors: Diane Beauvais and Andy Henson
National Physical Laboratory

ABSTRACT

This paper summarises the benchmarking of the Calibration and Measurement Capabilities (CMCs) as declared in the Bureau International des Poids et Mesures (BIPM) key comparisons database (KCDB), of a number of National Metrology Institutes (NMIs) including EUROMET countries and the USA, using validated data. This project was realised at the International Office, National Physical Laboratory for the UK Government Department of Trade and Industry. The benchmarking and the results are described in full, in the NPL report 'International Benchmarking and Analysis of UK Calibration & Measurement Capability - April 2002'.

1 BACKGROUND

The mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes [1] (the Mutual Recognition Arrangement or MRA) was drawn up by the intergovernmental 'Comité International des Poids et Mesures' (International Committee for Weights and Measures, CIPM). It was signed in October 1999, by the directors of National Metrology Institutes (NMIs) from 38 countries and 2 international organisations. The MRA is a response to the need for measurements and tests that are appropriate, reliable and trusted domestically and internationally in support of trade agreements, regulation and fair competition. It aims to reduce technical barriers to trade and leans towards the philosophy of "measured once, accepted everywhere". The MRA is subject to a transition period expected to be completed by the end of 2003.

The MRA is certainly the most significant development in the structure of metrology since the adoption of the 'Système International d'Unités' (International System of Units, SI) in the 1960's and arguably the major development since the signing in 1875 of the 'Convention du Mètre' (Metre Convention) in Paris by representatives of seventeen nations.

The MRA can be thought of as being supported by three pillars that require the participating NMIs to:

- Take part in appropriate international comparisons of measurements - **the key and supplementary comparisons**;
- Declare and subject their **Calibration and Measurement Capabilities (CMCs)** to extensive peer review;
- Implement and demonstrate an appropriate **quality system** (based on ISO 17025 in Europe).

One of the main aspects of the MRA is the BIPM key comparisons database (KCDB) available for public consultation on the BIPM website (www.bipm.fr).

The operational elements of the database are:

- Results of key and supplementary comparisons;
- Calibration and measurement capabilities (CMCs);
- Listed key and supplementary comparisons.

All CMCs undergo 100% peer review within the local Regional Metrology Organisation (RMO) and sample peer review by all other RMOs. They then have to be agreed by the Joint Committee of the RMOs and the BIPM (JCRB) before they can populate the database.

CMCs are or will be declared for the following 9 quantities:

- Acoustics, Ultrasound and Vibration
- Amount of Substance
- Electricity and Magnetism
- Ionising Radiation
- Length
- Mass and Related Quantities (including Flow)
- Photometry and Radiometry
- Thermometry
- Time and Frequency

CMCs are described using a common format, the so called ‘vocabulary list’ that has been agreed internationally. The vocabulary list provides consistency, facilitates the peer review and aids the end users. It is also fundamental to both the establishment of the database and the ability to benchmark the data contained in it.

Thus for the first time, the MRA provides end users with validated data declared in a harmonised way by NMIs that are obliged to operate a quality system and take part in international comparisons. End users are now able to make a realistic comparison of the services and uncertainties offered by the various NMIs and there are now data available in a format that is meaningful for a benchmarking study.

2 BENCHMARKING

The current benchmarking [2] builds on a similar unpublished study carried out by the authors in late 2000. At that time, none of the data had been through the full validation process. The study was informally shown to the directors of a number of key NMIs as part of the process of examining the possibility of increasing co-operation between the institutes. It generated sufficient interest to warrant repeating the process with up-to-date and validated data, thus enabling a report to be issued that could be published.

2.1 The benchmarking model

The current benchmarking considered only CMCs that had either completed the JCRB review (and are entered in the KCDB) or had at least completed the local RMO review stage (on the basis that experience from the previous study demonstrates that only minor changes in terms of benchmarking are likely to occur after the local RMO review).

At the time of the benchmarking the CMCs data for Acoustics, Ultrasound and Vibration, Length, Electricity and Magnetism, Amount of Substance and some of the data for Photometry and Radiometry had been through the review process, are now available on the KCDB and were benchmarked. The data for Ionising Radiation, Mass and Related Quantities and Thermometry had passed the first stage of the local RMO review and were also benchmarked. Whilst Flow and Time and Frequency had been included in the informal study with non-validated data, the data are still under discussion by the local RMO and thus for that reason these two quantities were not benchmarked in this study.

As resources were limited, for each quantity 10 countries were benchmarked (only 9 for ionising radiation), that is the UK, USA and a sample of eight countries from EUROMET. The countries remained constant within a quantity but varied from one quantity to another as appropriate. The countries from EUROMET included NMIs that offer State of the Art (SoA) capabilities but also a representation of other laboratories offering a range of services.

The CMCs data were rated per country by the UK EUROMET contact persons, strictly based on the data declared in the CMCs tables, at the 'line of service' level as described below in table 1.

Table 1. Vocabulary list, example for length.

Agreed vocabulary	**Divisions**
Length	Quantity
2 Linear Dimensions	Sub-quantity
2.1 Length Instruments	Service category
2.1.1 Interferometer	**Line of service**
CMC data entry	CMC data entry line
CMC data entry	CMC data entry line

The benchmarking model addressed two criteria:

- The range of services offered by each NMI was rated and compared to the range of services offered at the UK NMI, the so-called '**coverage'**. The following marking system as shown in table 2 was applied.

Table 2. Marking system for coverage.

0	Service not offered by this laboratory
1	Coverage significantly less than UK NMI
2	Coverage less than UK NMI
3	Coverage similar to UK NMI
4	Coverage greater than UK NMI
5	Coverage significantly greater than UK NMI

In the limited number of instances where the UK did not offer a line of service the benchmark was taken to be the mean of service coverage of those countries benchmarked that did offer the service.

- The **Performance** of each NMI in terms of the declared uncertainty was established as follows; the UK EUROMET contact person established the State of the Art (SoA) from the CMC data. The matrix below (Figure 1) was then used to judge the performance, taking into account that a number of CMC data entry lines might contribute to a single line of service for any given laboratory. The matrix was operated in the following way:

- For each country and each line of service, the best uncertainty declaration (data entry line level) was assessed in the range Poor to SoA;

- The same process was repeated for the remaining uncertainty declarations (data entry line level) contributing to this line of service, with the remaining uncertainty declarations considered together as a group;
- Using the matrix (Figure 1) the appropriate mark was allocated per line of service.

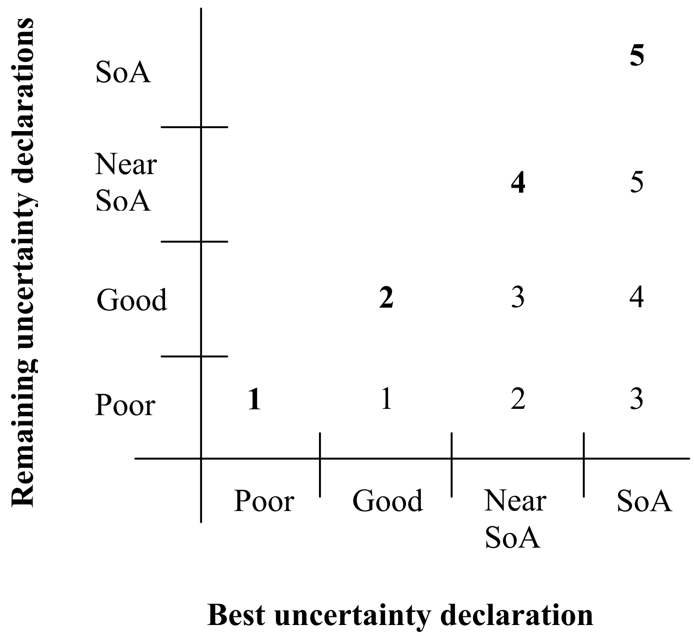

Figure 1. The performance matrix.

For each quantity, the data were collated per service line and per NMI. The data were then amalgamated in appropriate sub-quantities, following the vocabulary list and further consolidated to give an overview of the performance of each country per quantity. The results are shown as pictorial representations with the performance in terms of uncertainty level as the ordinate and the coverage as the abscissa at both sub-quantity and quantity level.

The exercise required the individual examination and comparison by the UK EUROMET contact persons of several thousand lines of data, the recording of nearly 10,000 numerical scores and the generation of more than 50 sub-quantity and quantity graphs.

To test impartiality in the assessment exercise, one of the leading non UK NMI was asked to independently repeat the benchmarking for a sample sub-quantity. Impedance was chosen (by that NMI), as it represents a particularly challenging field to judge with many parameters to take into account. This blind check showed that the judgements from the two NMIs were consistent within the accuracy of the study.

As part of the study, it was also considered appropriate to record the traceability arrangements, which are listed in the CMCs on the format that is available to the EUROMET contact persons.

2.2 Qualifications of the model and the data

The model was developed to be used across all the quantities and also to cope with the differences between the quantities and the CMCs declarations, but still requires some qualification:

- No rating of importance of the services was provided, i.e. no difference was made between critical and less important services. All services were assigned the same weighting factor (weighted scores were used in the original study but made little overall difference and did not warrant repeating);
- The judgement was made strictly on the CMCs statements, as they existed at the time of the benchmarking;
- The judgement was made by the UK EUROMET contact persons, with a small sample check conducted by another NMI;
- A graphical approach is used to represent the results in an easily digested format, however these representations should be considered as pictures rather than mathematical graphs;
- Due to the lack of validated data, Flow and Time and Frequency were not benchmarked;
- Only 2/3 of the CMCs finally expected to populate the KCDB are currently on the database. However approximately 90% of the European data have completed the EUROMET RMO review and were available for the benchmarking exercise.

3 RESULTS AND ANALYSIS

The benchmarking generated a wealth of data that is suitable for ‘data mining’ in many ways. Generally for detailed analysis the sub-quantity representations give more appropriate information, whilst the quantity representations are more suited to give a general overview only.

The key players and their relative performance can be identified by quantity and sub-quantity. The study confirms that the UK NMI is amongst the leading NMIs. It also demonstrates that smaller NMIs are able offer SoA services in the areas in which they specialise, with many taking traceability from other NMIs in the remaining areas. Nearly eighty traceability arrangements were recorded in the report, illustrating the fact that such arrangements do exist and that in appropriate circumstances devolution between NMIs can work.

It can also been seen that in general, laboratories are able to declare similar levels of uncertainty for some ‘classical’ quantities (such as Length, figure 2). Less mature quantities (such as Photometry and Radiometry, figure 3) tend to have a greater spread in their performance level. It seems likely that this is due to the rapid rate of development and the novelty of the measurements in these newer areas.

On the graphs each point represents an individual country.

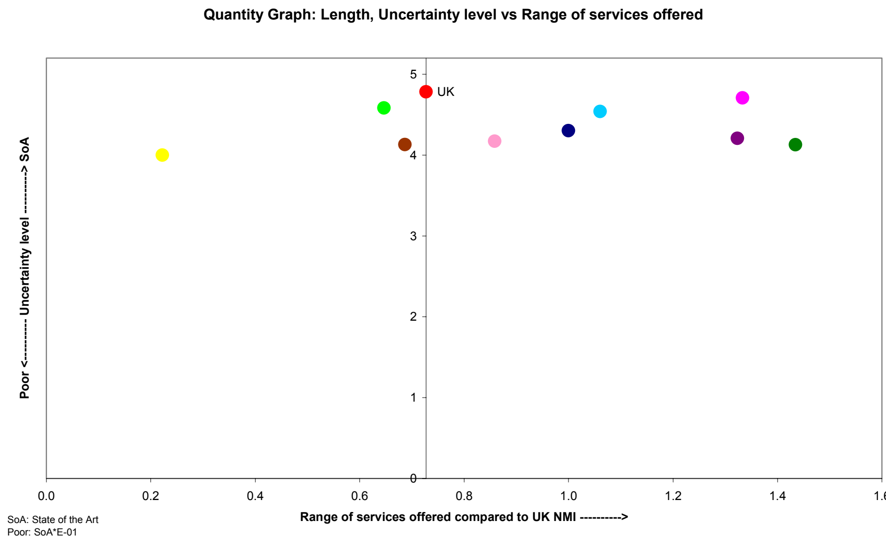

Figure 2. Quantity graph for Length.

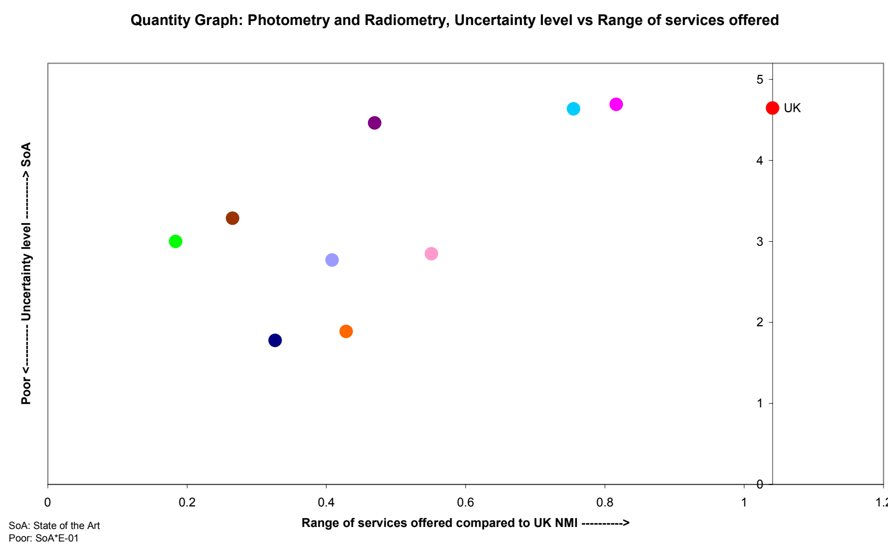

Figure 3. Quantity graph for Photometry and Radiometry.

The analysis of the benchmarking findings highlighted an issue regarding not so much the benchmarking but the rather the CMCs data. The 'Technical supplement to the MRA' describes the CMCs to be declared as those being '*ordinarily available to customers'* [1]. Whilst the MRA has achieved a very high level of harmonisation in the presentation of NMIs services, it is apparent that for a variety reasons NMIs have adopted a range of interpretations of this definition. In addition, some services considered as routine are available at the NMIs in one country and therefore included in the benchmarking, whilst in the case of a second country these services are devolved to accredited laboratories, in which case the services will not be included in the CMCs statements nor in the benchmarking although the services are available in that country.

Consequently, the range of services may appear to be under represented for a given country, and comparisons using the graphs must be duly qualified.

The benchmarking exercise demonstrated that comparison is possible, but also that interpreting the data is particularly challenging. This is illustrated by the sub-quantity ‘impedance’ where many parameters have to be taken into account. The ‘equivalence’ referred to in the MRA needs to be understood as the equivalence of the validity of the data. The quoted uncertainties do vary from NMI to NMI and end users may require assistance in using and exploiting the database.

4 CONCLUSION

The MRA has made publicly available, for the first time, the statements of Calibration and Measurement Capabilities that have undergone peer review, are supported by international comparisons and generated by NMIs operating a validated quality system. Progress during the MRA transition period appears to be on schedule with the KCDB containing 2/3 of the data finally expected to populate it. The data available on the KCDB are declared in a common format aiding the end users and allowing the benchmarking to be undertaken.

The NPL benchmarking report provides a data mine that can be used for detailed analysis at sub-quantity level. The study shows that some smaller NMIs are specialised in some areas offering SoA services and have devolved services in other areas, a route that larger NMIs might need to follow as new demands stretch resources.

The study also identifies that for some quantities many NMIs deliver services of essentially equivalent performance. This may be justified by the need for local service delivery. However the findings provide a useful context to discuss whether the scarce resources available at the highest levels of metrology are best utilised.

The MRA is clearly a major step forward but for many end users assistance will be necessary for them to be able to use the database directly. Providing this assistance is the role of the local NMI who use their knowledge and expertise to bridge the KCDB to the end users needs.

In a wider context, the issue raised by the benchmarking supports the idea, now being addressed in the EUROMET strategy, of closer collaboration between NMIs creating greater critical mass, possibly reducing unnecessary duplication of similar capabilities while respecting the need to provide local expertise to support end users.

5 REFERENCES

1. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes. See BIPM web site: www.bipm.fr.

2. D. Beauvais, A.S. Henson, International Benchmarking and Analysis of UK Calibration & Measurement Capability - April 2002, *NPL report INTOFF 1*, 2002.