IMEP Interlaboratory Comparisons – A Tool to Establish Confidence in Measurements

Ines Baer, Lutgart van Nevel*, Beatriz de la Calle, and Philip Taylor

Abstract: The International Measurement Evaluation Programme (IMEP) is owned by the European Commission's Joint Research Centre. IMEP organises interlaboratory comparison (ILC) in support of the policies of the European Union, covering a wide range of measurement tasks from food safety to environmental pollution. IMEP can be considered as a whole metrological process underlining the need for reliability and comparability of measurements. The article describes shortly the history and purpose of IMEP, and furthermore sums up its activities, links and impacts on the European measurement community.

Keywords: IMEP · Interlaboratory comparison · Measurement quality · Metrology

1. Introduction

The International Measurement Evaluation Programme (IMEP) organises interlaboratory comparisons (ILCs) in support to EU policies. IMEP is a registered trademark of the European Commission and is integrated in the Joint Research Centre (JRC) – Institute for Reference Materials and Measurements (IRMM).[1] Its ILC projects cover a wide range of measurement problems from food safety to environmental pollution.

The mission of the IRMM is the promotion of a common European measurement system, especially for internal market, environment, health and consumer protection standards. Measurements need to be reliable, traceable and comparable in order to support legislation and regulations, establish international trade conditions and to secure the quality and safety of food and feed. Thus laboratories need to be able to demonstrate that their measurement results are reliable and comparable. This also means that they must prove to be in compliance with legislation, international standards and international recognition arrangements that support the free trade goal 'measured once, accepted everywhere'. IMEP enables laboratories to assess their measurement performance and demonstrate their competence on a high quality level to accreditation, authorisation, and inspection bodies as well as to their customers.[2]

IRMM is one of the world's leading reference material producers, an expert adviser in food safety and quality and bio-analysis issues as well as a valued provider of reference measurement data. Presently, IRMM operates four Community Reference Laboratories (CRL). Its management system is certified according to ISO 9001, ISO 14001 and OHSAS 18001 and its units hold several accreditations.

IRMM not only demonstrates its measurement competence (ISO 17025 related activities) but also organises interlaboratory comparisons on behalf of BIPM (*e.g.* CCQM – *Comité consultatif pour la quantité de matière*) and EURAMET, thus offering the European laboratories benchmarking possibilities at the highest metrological level.

2. History of IMEP

The IMEP intercomparisons started in 1989 with a project about the measurement of lithium in serum.[3] The reference value to which the results were later compared was provided by IRMM using isotope dilution mass spectrometry (IDMS). The outcome of the exercise was that the outlier from the results had the closest agreement with the reference value and the other laboratories appeared to have a systematic measurement deviation. So, the main issue in the beginning of the programme was about rising awareness about measurement problems. In the meantime, IMEP has become a global equivalence demonstrator programme.

In the early days IMEP became known because it was one of the few intercomparisons where the assigned value was not based on the consensus value from the participants' results. And it was a general surprise when the first IMEPs showed an unexpected large spread of results, even for simple measurements, such as elemental contents in pure water. This led to the establishment of a questionnaire for each IMEP with the aim to determine the cause of this spread of results. Participants were asked to give information about their methodological approach, the traceability with regard to the reference material, the evaluation of uncertainty, and their expertise (their accreditation status and number of the respective type of analysis per year). The answers were evaluated following an univariant model, meaning it was verified if and how each one of these points contributed to the observed variation. Recently, IMEP has begun to approach this evaluation using multivariate analysis.

In the early nineties most of the ILCs were run in connection with health and environmental issues involving analysis of trace elements in serum or water. For the latter, comparisons of measurement were particularly relevant for member states, water being a more 'international product' than any other matrix considered by IMEP so far. Later on, ILCs started to focus more on the food & feed area, which is very much reflected by the rather recent start of the intercomparisons run in collaboration with the CRL for heavy metals in feed and

Correspondence: L. van Nevel
EC-JRC-Institute for Reference Materials and Measurements (IRMM)
Retieseweg 111
B-2440 Geel, Belgium
Tel.: + 32 14 57 1682
Fax: + 32 14 57 1685
E-mail: Lutgart.van-Nevel@ec.europa.eu

food.[4] All realised projects are shown in Table 1 and Table 2 – the corresponding reports can be found on the IMEP webpage.[5]

Table 1. Overview of all IMEP intercomparison since 1989

IMEP® Round	Title	Time Period
IMEP-1	Li in human serum	1989
IMEP-2	Cd in polyethylene	1990–91
IMEP-3	Trace elements in water	1991–93
IMEP-4	Trace elements in bovine serum	1991–95
IMEP-5	Fe in human serum	1991–94
IMEP-6	Trace elements in water	1994–95
IMEP-7	Trace elements in human serum	1997–98
IMEP-8	$n(^{13}C)/n(^{12}C)$ and $n(^{18}O)/n(^{16}O)$ in CO_2	1997–99
IMEP-9	Trace elements in water	1998–99
IMEP-10	Trace elements in polyethylene	1997–98
IMEP-11	Metals in car exhaust catalysts	1998–99
IMEP-12	Trace elements in water	2000–2001
IMEP-13	Trace elements in polyethylene	1999–2000
IMEP-14	Trace elements in sediment	1999–2000
IMEP-15	Trace elements in water	2001–2002
IMEP-16	Pb in wine	1999–2001
IMEP-17	Trace and minor constituents in human serum	2002–2003
IMEP-18	S in fuel	2004–2005
IMEP-19	Cd in rice	2002–2003
IMEP-20	Trace elements in tuna fish	2003–2004
IMEP-21	Trace elements sewage sludge	2005–2006
IMEP-22	Sulphur in petrol	2006–2007
IMEP-23	PAHs in water in presence of humic acids	2007–2008
IMEP-24	Heavy metals in toys	2008–2009
IMEP-25	Bromate in water	2009
IMEP-26	3-MCPD in bread	Ongoing
IMEP-27	Trace elements in mineral feed	2008
IMEP-28	Heavy metals in food supplements	2009

Table 2. Overview of all intercomparisons in collaboration with CRL for heavy metals in feed and food

CRL IMEPs	Heavy metals	Time Period
IMEP-101	In brown bread	2006
IMEP-102	In mineral water	2007
IMEP-103	In feed	2007–2008
IMEP-104	In seafood	2008
IMEP-105	In mineral feed	2008
IMEP-106	In food supplements	2009

3. What is IMEP?

IMEP supports EU policies by organising intercomparisons in the framework of specific EU Directives, or on request of a specific Directorate-General of the European Commission. The respective test material and analytes under investigation are always related to European legislation, preferably regulation or directives to be implemented in the Member states or in the process of revision. IMEP can be considered as a whole metrological process underlining the need of reliability and comparability of measurements. The ILCs are one-off exercises and hence address every time a different analyte/matrix combination, trying to get as close as possible to real-life samples. The aim of this is to give an overview about the actual measurement capability status in various fields where the EU policies are relevant.

IMEP provides support to the European measurement infrastructure in the following ways:
- IMEP *distributes metrology* from the highest level down to the field laboratories. These laboratories can benchmark their measurement results against the IMEP certified reference value. This value is obtained independently from the participants' results and is established according to metrologically best practice.
- IMEP helps laboratories to assess their estimate of *measurement uncertainty*. The participants are invited to report the uncertainty on their measurement result. IMEP integrates the estimate into the scoring, and provides assistance for the interpretation.

The traceability of IRMM's reference values to the SI and frequently the demonstrated equivalence of its measurement results to those of BIPM and the national metrology institutes world-wide underpin its acceptance as provider of interlaboratory comparisons to the field laboratories.

IMEP fulfils the requirements described in ISO/IEC 17025:2005 and in ISO guide 43:1.

3.1 *Establishment of the Certified Reference Value (CRV)*

The idea behind the use of a CRV is that the best possible values are used as reference and that they are obtained from well-understood measurements processes rather than simply *via* a consensus approach. They are required to demonstrate traceability and should have a very small uncertainty, evaluated according to international guidelines. IMEP obtains its CRV thanks to a network of renowned institutes which have proven measurement capabilities. They are National Metrology Institutes (NMI) or well-known expert laboratories in the respective specific field. These laboratories prove their measurement capability in various ways, *e.g.* by participating in international, mutually recognised CCQM comparisons, or internationally recognised publications, successful participation in proficiency tests, *etc*.[6]

Consequently, the methods used for the determination of the CRV are, if possible, primary methods of measurements (*e.g.* IDMS), but other methods are also used for control purposes, and in some cases internationally agreed measurement methods might be found to be suitable. It is important that the uncertainty obtained *via* the reference method is small enough to be fit for purpose. The establishment of the CRV and the measurement capabilities of the reference institutes are published in a separated certification report, or included in the participants report.

To give an example, in IMEP-23 the parameters were the eight priority polycyclic hydrocarbons (PAH) as defined in the EU Water Framework Directive.[7] The matrix in this ILC was ground water with added humic acid. Certification measurements were carried out by two experienced reference labs, but their measurement results appeared to differ for the large congeners (Fig. 1). It appeared that both labs had used a deuterated internal standard, which was added 24 h before analysis by lab 1 whereas lab 2 added it directly before analysis.[8] Just like the PAHs, the internal standard (IS) needed time to adsorb to the humic acid. So, if an IS is added just before analysis and given insufficient time to reach the adsorption equilibrium, its recovery is higher than the recovery of the sample and this leads to an underestimation of the measurand. As result, the values from lab1 were given priority over the lab 2 values for establishing the reference values.

3.2 *Measurement Uncertainty*

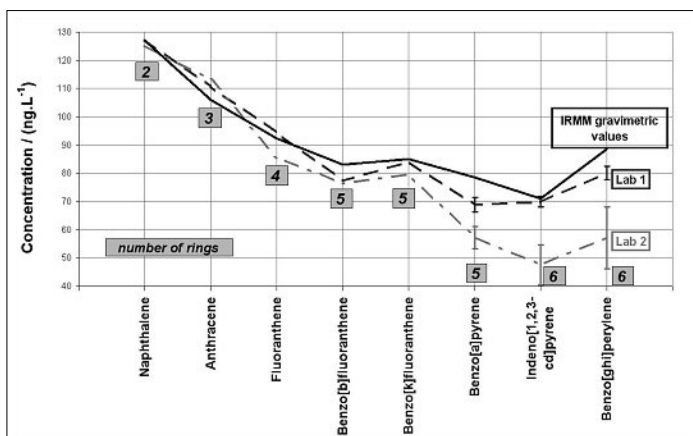In an IMEP ILC, laboratories are asked to report not only a value but also its asso-

Fig. 1. Results for the eight PAHs from the two certifying laboratories in IMEP-23.

ciated uncertainty and the expansion factor (expanded uncertainty). No results are excluded and they are represented in relation to the reference material's uncertainty (u_{ref}) and the proficiency test standard deviation (σ). The individual laboratory performance is expressed in terms of z (Eqn. 1) and zeta scores (Eqn. 2) in accordance with ISO 13528 and the IUPAC International Harmonised Protocol:[9,10]

$$z = \frac{x_{lab} - X_{ref}}{\sigma} \qquad (1)$$

$$zeta = \frac{x_{lab} - X_{ref}}{\sqrt{u_{ref}^2 + u_{lab}^2}} \qquad (2)$$

z = z score: deviation in units of σ; $zeta$ = zeta score: deviation in units of composed uncertainty of laboratory and the reference value; x_{lab} = measurement result reported by a participant; X_{ref} = certified reference value (assigned value); u_{ref} = standard uncertainty of the reference value; u_{lab} = standard uncertainty reported by a participant; σ = standard deviation for proficiency assessment.

Both scores (Eqn. (1) and Eqn. (2)) can be interpreted as: satisfactory result for |score| ≤ 2, questionable result for $2 <$ |score| ≤ 3 and unsatisfactory result for |score| > 3. The *z score* indicates whether a laboratory is able to perform the measurement in accordance with legislation, state-of-the-art, experts opinion, *etc.*

The *zeta score* provides an indication of whether the estimate of uncertainty is consistent with the laboratory's deviation from the reference value.[10] An unsatisfactory zeta score may be caused by an underestimated uncertainty or by a large deviation from the reference value, or to a combination of the two factors. A laboratory with an unsatisfactory zeta score has an estimation of the uncertainty of its measurements which is not consistent with laboratory's deviation from the reference value. The standard uncertainty should fall in a range between a minimal required

(u_{min}), and a maximal allowed (u_{max}) reported standard uncertainty. u_{min} is set to the standard uncertainty of the reference value. It is unlikely that a laboratory carrying the analysis on a routine basis is able to measure the measurand with a smaller uncertainty than the reference laboratory itself. u_{max} is set to the standard deviation accepted for the proficiency test, σ. If the standard uncertainty from the laboratory, $u_{lab} < u_{min}$ it is likely that the laboratory has underestimated its uncertainty. If $u_{lab} > u_{max}$, some effort should be made to reduce it because it exceeds the present state-of-the-art in that field of analysis. If zeta > 2, the results disagree within the expanded uncertainties, if the *k*-factor is chosen so that the expanded uncertainty is 95%.

The standard uncertainty for values of the laboratory is calculated individually by dividing the reported expanded uncertainty by the reported coverage factor. The procedure followed in case of non-reporting depends on each project, but is based on international guidelines.[9,10]

4. Practical Examples of IMEP Comparisons

4.1 *IMEP-20: Trace Elements in Tuna Fish*

Fish and fishery products are a central element in the diet of a good part of the world population. Some years ago, concerns rose about health risks due to elevated contents of mercury in fish. To protect public health the relevant regulations set maximum levels for Hg and Pb in fishery products.[11,12] In this context, an IMEP exercise was organised to verify the relevant laboratories' capability to measure the questioned trace element. IMEP-20 was organised in collaboration with the Community Reference Laboratory for Residues – Istituto Superiore di Sanità, Rome (CRL-ISS) in support to the European network of National Reference Laboratories (NRLs).

Participants in IMEP-20 were offered to measure the total amount content of As,

Hg, Pb, Se and methylmercury (CH_3Hg^+) in tuna fish. Methylmercury-cation was included in this exercise because it is the main source of human intake of Hg. However, only a few laboratories reported results and its determination is still an issue that arises regularly. Arsenic is also highly toxic and rules for measurements of As are set in the commission decision on implementing council directive 96/23/EC concerning the performance of analytical methods and the interpretation of results regarding residues.[13] Selenium was also included because within a narrow range it is both essential and toxic and it has been associated with certain forms of cancer and cardiovascular diseases.

The tuna fish sample used as test material in IMEP-20 had been withdrawn from the market due to its elevated Hg content. It was perfectly appropriate for this IMEP exercise because of its homogenous distribution of contaminants. The reference measurements for Hg and Pb were carried out using ICP-ID-MS.[14] Those for As and Se were performed by k_0-neutron activation analysis (NAA).[15] The certified value for methylmercury was established *via* species-specific gas chromatography (GC) ID-MS.[16] Measurement results were reported by 235 participants from 14 EU member states, 10 acceding countries and 3 candidate countries. Amongst those there were 23 NRLs and 38 laboratories nominated by a NRL. In the frame of the collaboration between IRMM and the European Co-operation for Accreditation, 61 laboratories were nominated *via* their National Accreditation Bodies (NABs) and reported measurement results in IMEP-20. In Fig. 2 the overall results for one trace element is shown.

4.2 *IMEP-23: The Eight WFD PAHs in Water in the Presence of Humic Acid*

As mentioned earlier, IMEP-23 studied the capability of analytical laboratories to measure total concentrations of the eight WFD (EU water framework directive) PAHs in the presence of humic acid in a water matrix. These include seven polycyclic aromatic hydrocarbons (PAHs): naphthalene, anthracene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]pyrene, indeno[1,2,3-cd]pyrene and benzo[ghi]perylene, plus fluoranthene as an indicator substance.

The 59 participants were invited *via* the IMEP Regional Coordinators, the IRMM website, the European Co-operation for Accreditation, the International Committee for Protection of the Danube River and the International Committee for Protection of the Rhine. For the reported results z scores were calculated with a target standard deviation of 20% of the reference val-
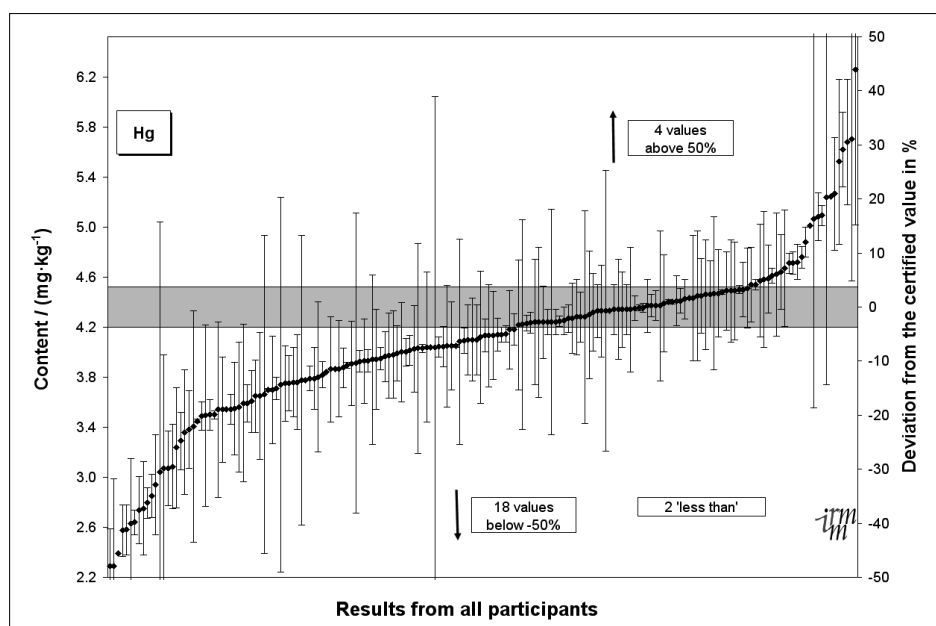
Fig. 2. Results for Hg from IMEP-20: Trace elements in tuna fish.

5. IMEP and the International Structured Measurement System

IMEP assists the establishment of an internationally structured measurement system for chemical measurements and thus interacts with several international institutions.[3] The European Cooperation for Accreditation (EA), the International Federation of Clinical Chemistry (IFCC), the WMO, the International Organisation of Wine and Vineyards (OIV), the European Committee for Standardisation, and the Forum of European Geological Surveys (FOREGS – Geochemical Baseline Programme) have used IMEP as a tool to benchmark laboratory performance.

Some examples of collaboration are described here:

i) IMEP assists the European accreditation in the monitoring of the EA multilateral agreement (MLA) and serves as a special tool for the National Accreditation Bodies (NAB) to ensure compliance of their accredited laboratories with the ISO/IEC 17025 Standard. The NABs nominate laboratories to participate in the relevant IMEP exercises to evaluate their performance against independent reliable reference values. Unsatisfactory results in IMEP have a direct consequence for EA-nominated laboratories – they are required to take immediate corrective measures. In this way, IRMM supports the improvement of the efficiency of accreditation in chemistry.

ii) Another aspect of the international collaboration is IMEP's assistance to the CIPM Mutual Recognition Agreement. The CIPM-MRA is a transparent process to underpin measurement capability claims of signatories to the meter convention. So called key comparisons are organised for NMIs or des-

ue. In addition, zeta scores were calculated for those participants who had reported an uncertainty estimate.

Humic acid was added as a simulation of natural colloidal organic matter and is known to adsorb PAHs. It was explained in a previous section (3.1) that this requires timely addition of a standard and if the IS is given insufficient time to reach the adsorption equilibrium, its recovery is higher than the recovery of the sample and this leads to an underestimation of the measurands. For this reason laboratories were given a specification for sample preparation together with the samples. Just like for the establishment of the reference value, the distribution of the results is not symmetric: the lower concentrations outweigh the higher and this for all congeners, however this effect is more

pronounced for the larger compounds (Fig. 3 left). This tendency is probably a direct consequence of participants underestimating the effect of adsorption onto the humic acid.

Kernel plots allowed the observation of two different distributions stemming from the two methods of analysis used.[17] Approximately half of the participants used HPLC with fluorescence detector, the other half GC/MS as the method of analysis. Measurement results obtained with HPLC were frequently characterised by a higher median than those obtained with GC. These data are visualised for indeno[1,2,3-cd]pyrene in Fig. 3 (right). The same trend was observed for all congeners. This may be subject of further study outside the frame of this interlaboratory comparison.
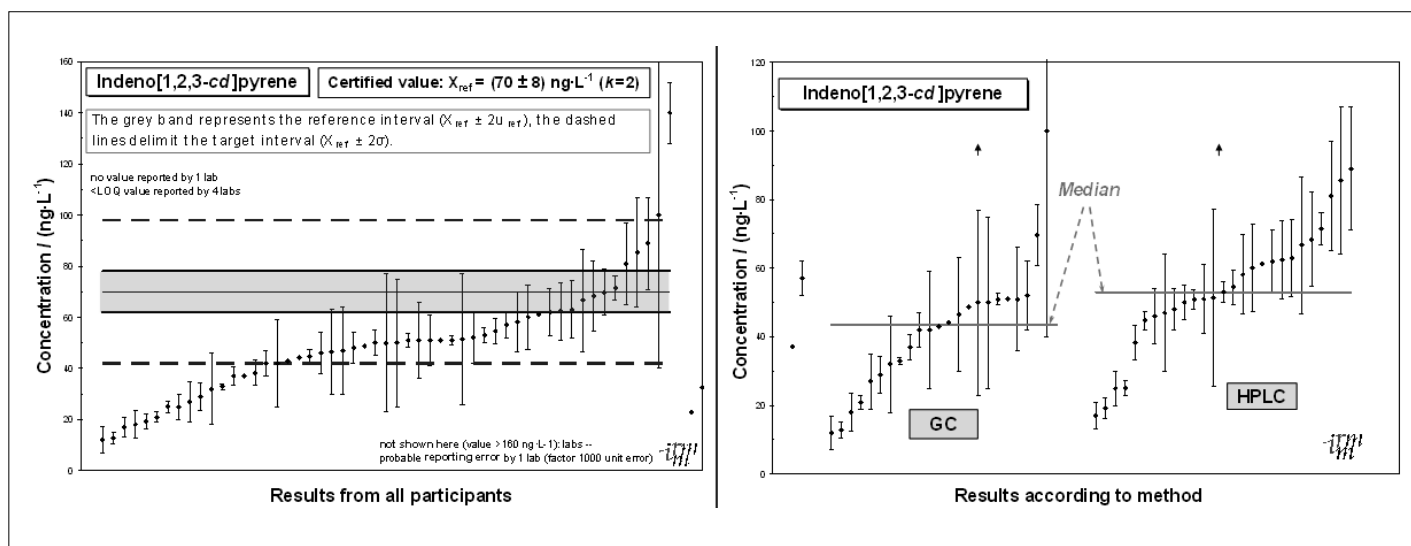

Fig. 3. Results for indeno[1,2,3-cd]pyrene from IMEP-23: PAHs in water.
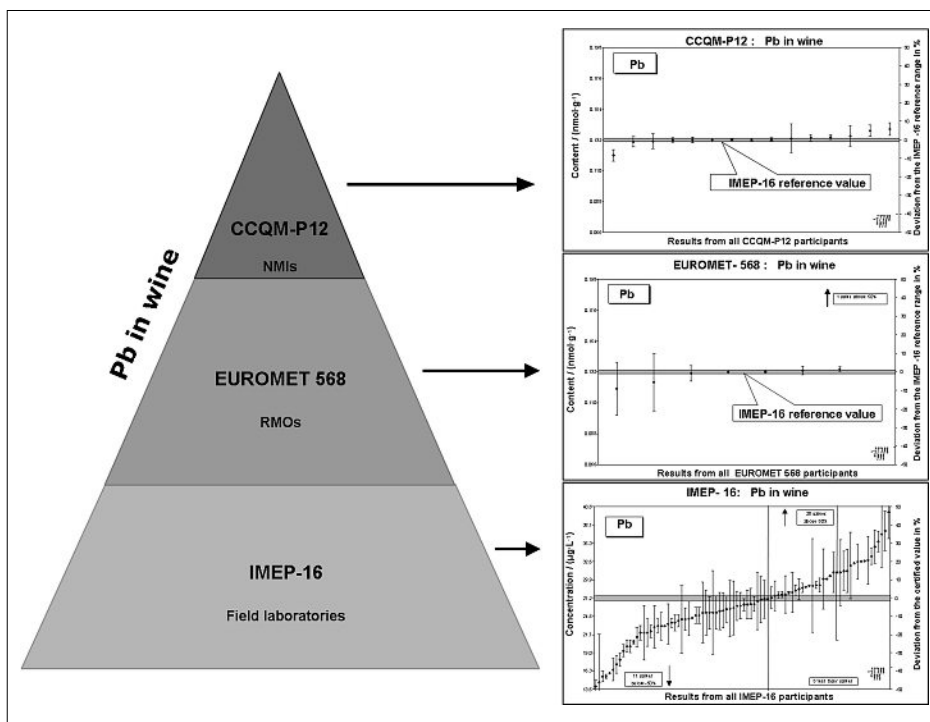
Fig. 4. Comparison of international measurement capabilities with a sample from IMEP-16: lead in wine.

[1] http://irmm.jrc.ec.europa.eu.
[2] Y. Aregbe, L. van Nevel, P. D. P. Taylor, Trace Elements – Institute for UNESCO, 2006, 11, 1.
[3] Y. Aregbe, T. Prohaska, P. Robouch, in 'The Determination of Chemical Elements in Food: Applications for Atomic and Mass Spectrometry', Ed. S. Caroli, Wiley & Sons, New York, 2007, p. 167.
[4] http://irmm.jrc.ec.europa.eu/html/CRLs/crl_heavy_metals/index.htm.
[5] http://irmm.jrc.ec.europa.eu/html/interlaboratory_comparisons.
[6] http://www.bipm.org/en/committees/cc/ccqm/.
[7] Commission Directive 2000/60/EC, Ref L 327/1, 2000.
[8] J. van de Kreeke, B. de la Calle, S. Bynens, I. Verbist, P. Taylor, O. Bercaru, M. Ricci, B. Sejeroe-Olsen, P. Shegunova, A. Bau, Interlaboratory comparison report, Report EUR 23287 EN, Belgium, 2008.
[9] ISO 13528: 2005, 'Statistical methods for use in proficiency testing by interlaboratory comparisons'.
[10] M. Thompson, M. L. R. Ellison, R. Wood, Pure Appl. Chem. 2006, 78, 145.
[11] Commission Regulation (EC) No. 466/2001, Ref L 77/1, 2001.
[12] Commission Directive 2001/22/EC, Ref L 77/14, 2001.
[13] Commission Decision No 2002/657/EC, Ref L 221/8, 2002.
[14] J. P. Snell, C. R. Quétel, J. Anal. Atom. Spectrom. 2005, 20, 447.
[15] Y. Aregbe, C. Quétel, P. D. P. Taylor, Metrologia, Tech. Suppl. 2004, 41, 08004.
[16] J. P. Snell, C. R. Quétel, L. Lambertson, J. Qvarnstrom, J. Anal. Atom. Spectrom. 2004, 19, 1315.
[17] AMC Technical Brief, 2006, AMCTB No 4; http://www.rsc.org/amc/.
[18] http://irmm.jrc.ec.europa.eu/html/CRLs/crl_heavy_metals/interlaboratory_comparisons/index.htm.

ignated laboratories to demonstrate the equivalence of their measurement in view of mutual acceptance of national standards and reference measurements. Like in a normal IMEP comparison, the institutes are free to use the procedure and method of their choice, but in order to support their measurement capability claims they need to report reliable results of highest metrological quality with demonstrated traceability and a complete uncertainty budget.

The IRMM also coordinates ILCs in support to institutes which take part in regional metrology organisations like EURAMET (European Association of National Metrology Institutes) or APMP (Asia Pacific Metrology Programme).

On occasion, ILCs could be run in parallel for the different metrological levels. Meaning that the same sample material could be used for key comparisons for National Metrology Institutes, for an ILC for the Regional Metrological Organisations (RMO) such as EUR-AMET, and finally for regular field laboratories via IMEP. An example of such a case is represented in Fig. 4 where the results from a key comparison *Pb in wine* organised by BIPM/CCQM are shown at the top of the international measurement infrastructure, followed underneath by the results of the EUR-AMET trial and finally the overall participants' results of the corresponding IMEP-16. For comparison reasons, the IMEP-16 reference value is introduced in the three graphs. The increase of the spread of the results while going down the measurement structure is very well represented here. Another advantage of these comparisons is that a regular field laboratory can compare its results with its NMI.

iii) Finally, IMEP supports the work of the CRLs, which were created to implement Regulation (EC) No 882/2004 on official controls performed to ensure the verification of compliance with the feed and food law, animal health and animal welfare rules. The CRL for heavy metals in feed and food, for instance, is a regular customer of IMEP. It is established at the IRMM, and works together with appointed national reference laboratories (NRLs) of the EU Member States. To reduce the risk to human health associated with high heavy metal content in food and feed, maximum allowed limits in several commodities have been laid down in the European legislation. The CRL acts in this frame of legislation dealing with controls of heavy metal content in food and feed as a result of environmental or industrial contamination. One of the core tasks of the CRL is to organise interlaboratory comparisons specifically for appointed NRLs. And IMEP regularly organises these ILCs for the CRL for heavy metals in feed and food.[18]