CCQM Online Workshop on Digital and FAIR Chemical and Biological Reference Data and Certificates: Challenges and Opportunities

9-12 September 2024, virtual

Summary Meeting Report

Workshop Overview and Report Structure

The online workshop, held over 4 days, on Digital and FAIR Chemical and Biological Reference Data and Certificates was organized by the CCQM and hosted by the BIPM and brought together experts to discuss three main themes: challenges with unique interoperable identifiers for chemical and biological measurements, digitalization of CRM certificates, and best practices in developing databases for Chem/Bio data that meet FAIR (Findable, Accessible, Interoperable, and Reusable) principles.

All presentations by speakers were pre-recorded and are available for viewing via the workshop website at https://ccqmws2024.org/ . Each session was planned with outputs to be achieved and the areas for which recommendations should be developed through discussions at the workshop. Key topics for which the workshop wished to develop recommendations were:

a) What system(s) of unique interoperable identifiers should be used in both the BIPM KCDB and JCTLM databases;
b) Options for National metrology Institutes (NMIs) wishing to implement Digital CRM certificates and adhere to FAIR principles;
c) Key issues to be considered in database development and maintenance when wishing to apply FAIR principles.

The report below provides a summary of the topics presented and key points discussed during each day of the workshop. A more detailed description of discussions and recommendations of each day of the workshop can be found in the annex of the report.

Summary and outcomes of Day 1: Unique interoperable identifiers

The first day of the workshop commenced with a welcome address by Elvar Theodorsson (Linköping University), who emphasized the importance of digital and FAIR chemical and biological reference data. The primary objective was to explore the challenges and opportunities associated with creating unique interoperable identifiers in the chem-bio domain and specifically for their incorporation into databases maintained by the BIPM, notably the KCDB and JCTLM DB.

Stéphanie Maniguet presented on BIPM's databases – the Key Comparison Database (KCDB) and the Joint Committee for Traceability in Laboratory Medicine (JCTLM) database of reference materials. She highlighted the necessity for unique interoperable identifiers for calibration and measurement capabilities (CMC) and reference materials. Björn-Erik Erlandsson discussed the importance of standards like the Unified Code for Units of Measure (UCUM) in e-health and health informatics, emphasizing the need for a unified approach to quantities and units. Moulham Alsuleman from the

National Physical Laboratory underscored the importance of collaborative efforts to achieve digital transformation in research laboratories. Young Bae Hansen introduced the Nomenclature for Properties and Units (NPU) terminology for laboratory medicine, discussing the challenges in maintaining a stable and unambiguous terminology system. Richard Hartshorn explained the International Chemical Identifier (InChI) and InChI Key systems, highlighting their utility in linking chemical structures to databases. Janet Miles discussed the SI Digital Framework and the importance of globally accepted digital standards for metrology. Jeff Shick presented the Global Substance Registration System (GSRS) for pharmaceutical substances, emphasizing the importance of a common identifier for all substances used in medicinal products. Philip Strömert discussed the use of ontologies in chemistry, particularly within the NFDI4Chem initiative.

The participants identified several challenges in creating unique identifiers for diverse chemical and biological substances. Current efforts of creating unique interoperable chem/bio identifiers focus on the individual substances or their mixtures. Considerations must be made to accommodate the identification of sample matrix in a flexible manner that can accommodate various types of data. The importance of adopting existing standards and ontologies to avoid reinventing the wheel was highlighted.

A consensus opinion was to use InChI unique identifiers for measurands in reference materials and to explore InChI identifiers for measurands in secondary reference materials combined with newly developed identifiers for sample matrix, by leveraging the NPU naming and coding systems for the 'system' names and codes as an initial approach. The necessity of forming a working group to address these challenges was a key outcome of the workshop. The group would be tasked with developing a system for unique identifiers in chemical and biological reference data, leveraging existing standards and technologies, and would include experts from various fields and organizations to ensure a comprehensive approach.

Day 2: Digitalization of CRM certificates

The second day began with a welcome address by Adriaan van der Veen (VSL), who served as the moderator for the session. The focus of the day was on the digitalization of reference material certificates, exploring the challenges and opportunities associated with this process.

Dinis Camara from NIST presented the development of a digital reference material certificate (DRMC). He emphasized the need to model DRMCs independently of digital calibration certificates due to the unique requirements of reference materials. Michael Melzer from BAM discussed the development of digital reference material certificates derived from digital calibration certificates, introducing the utility model for digital documents and emphasizing the importance of machine readability and interoperability. Juris Meija from NRC Canada provided an overview of NRC's efforts in making machine-readable certificates, highlighting the benefits of digital calibration certificates, including improved discoverability and interoperability. Stephen Ellison from LGC focused on the digital representation of measurement uncertainty, discussing the various uses of measurement uncertainty and the need for digital representations that support subsequent evaluations. Mark Greiner from Merck KGaA discussed digital CRMs and blockchain-enabled certificates for the cannabis industry, introducing the M-Trust patent family and the Chemist Twin project. Toru Miura from Fujifilm Wako provided an overview of Fujifilm's experiences with certificate digitization, highlighting the advantages

and challenges of digital certificates. Xingchuang Xiong from NIM China presented the recent progress of digitalization for metrology in NIM.

The workshop identified several challenges and opportunities in the digitalization of reference material certificates starting from the fact that the resources and infrastructure required for digitalization cannot be taken for granted. Key points of discussion included the understanding of user expectations and demonstrating the benefits of digitalization, the ability to conduct audit of DRMCs, aspects surrounding the data security and the ability to protect the data associated with DRMCs, as well as discussions surrounding the extent of data that is subject of digitalization. The importance of engaging with users and stakeholders to gather feedback and ensure that digital certificates meet their requirements was emphasized. A discussion also took place on dealing with modifications of DRMCs, their discontinuance, and how it should be done in adherence with the requirements of ISO standards.

Day 3: Databases and FAIR data principles

The third day commenced with a welcome address by Carlos Gonzalez (NIST). The primary focus of the third day was on the application of FAIR principles to reference data and databases that host them.

Carlos Gonzalez emphasized the importance of data accessibility, universal data formats, and metadata standards. He highlighted the challenges in achieving universal standards across different communities, noting that it often requires significant social engineering to reach consensus. The need for flexible and extensible ontologies and metadata standards was also discussed. Additionally, the discussion touched on data security and intellectual property issues.

Sunghwan Kim from NIH provided an overview of PubChem, a public chemical information resource. He discussed the strategies employed to make data FAIR, including the use of persistent identifiers, support for programmatic access, data provenance, and machine-readable data. Stuart Chalk from the University of North Florida presented on the IUPAC Gold Book, detailing the efforts to update and digitize the Gold Book, including the assignment of DOIs to individual concepts and the development of an API for data access. Bob Hanisch from NIST discussed the attributes of FAIR data repositories, emphasizing the need for institutional commitment and data born FAIR. Patrick Hodapp from the Karlsruhe Institute of Technology introduced the Chemotion, an electronic laboratory notebook designed for recording and organizing experimental data. Carsten Kettner from Beilstein Institute gave an overview of the Strenda database that has been developed for enzymology data and which makes use of other databases such as PubChem. Stephanie Maniguet from BIPM provided an overview of the KCDB and its role in supporting the CIPM MRA. John Mund from the University of Colorado Boulder presented a project focused on creating a database for greenhouse gas measurements. Ben Place from NIST discussed the development of the DIMSpec database for mass spectrometry data and Adriaan van der Veen from VSL presented on the development over the years of the database being used to manage their large suite of primary reference gas mixtures (PRGMs). The presentation also pointed out the lack of adherence of some of the FAIR principles such as the use of unique identifiers.

The collection of examples of data repositories demonstrated how the FAIR data principles can be applied to data management and highlighted some of the limitations and issues for discussion during the workshop, most notably, the significant challenges in achieving interoperability across different research domains. The importance of ontologies, standard vocabularies, and metadata schemas was

emphasized, as well as the need for data stewards with domain knowledge and technical expertise to ensure the success of FAIR data initiatives. The discussion on infrastructure focused on the considerations for on-premise versus cloud storage, with an emphasis on IT security and data integrity.

Day 4: Summary and Recommendations

The last day of the workshop commenced with Carlos Gonzalez (NIST) welcoming participants. The primary focus of the day was on summarizing the technical discussions from the previous three days and formulating recommendations.

The session on unique interoperable identifiers emphasized the need for identifiers that are unambiguous, stable, and traceable, while avoiding redundancies. A clear distinction was made between nomenclature and terminology. It was concluded that a single system such as the InChI is insufficient to cover sample matrices, thus calling for the need of a set of ontologies. T

The session on the digitalization of CRM certificates focused on the need for harmonization of approaches to digital reference material certificates. Data security issues and opportunities for validating digital certificates was discussed, with recommendations to address these issues and establish minimum requirements that also address the plans for the long-term maintenance and access to the data. The infrastructure for data storage and management should be secure and reliable.

The session on databases and FAIR data principles highlighted the importance of harmonized terms and ontologies for chemistry and biology. Collaboration with organizations such as IUPAC and ISO should be pursued and more awareness needs to be placed to the existing ontologies and controlled vocabularies. Knowledge transfer on best practice for evolving to the use of primary databases for maintaining measurement results as an initial step in the FAIR data process was discussed.

The workshop concluded by summarizing the key recommendation from the previous 3 sessions.

Summary Recommendations

a) Recommendations related to what system(s) of unique interoperable identifiers should be used in both the BIPM KCDB and JCTLM databases:

- **A dedicated group of experts should be established:** to develop a system for unique identifiers in chemical and biological reference data with a focus on the sample matrices and needs for large and complex molecules, as well as identifiers for traceability and commutability. This group should include representatives from NMIs, external experts, and stakeholders from various fields.
- **Adopt and Adapt the Existing Standards:** Use InChI unique identifiers for measurands is recommended, whenever possible. The suitability of existing standards such as InChI, NPU, and LOINC for different chemical and biological substances should be assessed. These standards should be made interoperable and compatible with the SI Digital Framework, and guidelines should be developed for their consistent use across different platforms.
- **Focus on Interoperability:** Tools and systems that facilitate the interoperability of different databases and identifiers should be developed. This includes creating machine-readable

identifiers that can be used across various platforms, developing APIs and other tools to enable seamless data exchange between different systems.

b) Recommendations related to options for National metrology Institutes (NMIs) wishing to implement Digital CRM certificates and adhere to FAIR principles:

- **A dedicated group of experts should be established**: to address harmonization of approaches to the production of DRMCs. The resources to initiate and maintain a RM certificate digitalization effort need to recognized, and a roadmap developed on how the production of DRMCs can become accessible to NMIs that have resources to implement but not develop their individual solutions.
- **Understand User Expectations:** Users and stakeholders should be engaged to gather feedback and ensure that DCRMs meet their expectations and requirements. Open and standardized data formats were identified as key issues in digitalization efforts and will require dialogue and agreements with instrument vendors and regulatory agencies.
- **Ensure Data Security:** Security measures should be implemented to ensure the authenticity and integrity of digital calibration certificates. Digital signatures, encryption, and blockchain should be used to secure digital certificates. Guidelines should be developed on how digital certificates can be validated.
- **Consider Sustainability:** Digitalization efforts should be sustainable by choosing future-proof technologies and formats. A plan should be developed for the long-term maintenance and access to digital certificates.

c) Recommendations related to key issues to be considered in database development and maintenance when wishing to apply FAIR principles.

- **A group of experts is recommended:** to investigate the application of FAIR data principles to the KCDB and JCTLM databases.
- **Resource Compilation:** Resources and best practices for data management and FAIR data principles should be compiled and shared among NMIs.
- **Policy Development:** CCQM should consider its policy on making Key Comparison data publicly available in a machine-readable format. Implementation of such a policy would be expected to require developing templates and guidelines for consistent data reporting, as well as training.

Annex: Summaries of discussions of individual days of the workshop

Day 1 discussions: Challenges with unique interoperable identifiers in the Chem/Bio area

Issues which the workshop wished to focus on in Day 1

1. Gain understanding of the various **options for unique interoperable identifiers for the measurands** (within the scope of CCQM) that can cover both the component and matrix within the measurand description.
2. Gain understanding on **how to integrate** unique interoperable identifiers with the concept of broad-scope CMCs.
3. Review the structure and use CMC service categories for Chemical/Biological CMCs for the potential of **implementing FAIR principles** and relationship to the JCTLM and other databases.
4. **Develop recommendations** of what system(s) of unique interoperable identifiers should be used in both the KCDB and JCTLM databases (beyond the use of InCHI).

Introduction to discussions

The traceability and equivalence of laboratory results in chemistry and biology ultimately rest on traceability to international and national reference systems (1) based on reference materials and reference measurement procedures.

When computer systems in chemistry and biology were created in the 1960s, the need for naming and coding systems for measurands became obvious for reporting results and reimbursements. Such or similar identifiers could be used to identify reference materials in chemistry and biology, which are among the topics of the current CCQM workshop.

Currently used unique interoperable identifiers for measurands in the Chem/Bio area have originated in analytical chemistry or laboratory medicine.

Identifiers originating in analytical chemistry

The Chemical Abstract Service (CAS)

A CAS registry number is a proprietary, unique identification number assigned to all known chemical substances that the Chemical Abstract Service in the U.S.A. assigns to index the CAS registry (https://www.cas.org/cas-data/cas-registry).

The International Chemical Identifier (InChI)

InChI was developed jointly by the IUPAC and the National Institute of Standards and Technology (NIST) (2). Its purpose is to serve as a standard for encoding molecular information, optimized for searching for molecular information on the internet and in

databases. Its cornerstone is the highly formalized IUPAC chemical nomenclature. InChI is free and non-proprietary, developed and maintained by the InChI Trust (https://www.inchi-trust.org/iupac/).

Identifiers originating in Laboratory medicine

Snomed CT

Snomed clinical terms (CT) (https://www.snomed.org) is a proprietary computer-processable collection of medical terms and codes in all areas of medicine, including laboratory medicine (3). It represents the most comprehensive clinical healthcare terminology in the world.

LOINC

LOINC (https://loinc.org) is a widely used international naming and coding system for clinical observations and results in laboratory medicine owned and developed by the Regenstrief Institute (4). It has incorporated significant aspects of other systems for identification.

C-NPU

The IUPAC-IFCC NPU naming and coding system is a collaborative project between the International Union of Pure and Applied Chemistry (IUPAC) and the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) (5). It is free, non-proprietary, and unique in naming and coding measurands (quantities intended to be measured) in patients rather than in patient samples, as is the rule for other identifiers (https://npu-terminology.org). It is primarily used in the Nordic countries and the Czech Republic.

Identifiers for references in chemistry and biology

Identifiers for measurands in chemistry and biology must be unambiguous and defined in international consensus using scientific logic and proper ontological principles. An organization and sufficient resources over extended periods are needed to maintain and develop the system.

A measurand in pure ***primary reference material*** in chemistry and biology is optimally identified using **InChI** identifiers. CAS identifiers are proprietary and do not encompass the scientific rigor provided by the InChI system. Snomed CT, Loinc, and c-NPU systems are for use in medicine and laboratory medicine and do not provide the proper broad context and characterization of specific properties needed for reference materials in chemistry and biology.

Interoperable identifiers for secondary reference materials in chemistry and biology are for measurands/molecules stored in vitro in an appropriate matrix (6-8). Therefore, using separate identifiers for the measurand and the matrix is optimal. It is insufficient to identify *a measurand in* ***a secondary reference material*** for chemistry and biology using only an InChI identifier since the sample matrix influences the measurement results from numerous and commonly used non-selective measurement procedures and measuring systems.

Identifiers for sample matrices *need to be developed for secondary reference materials*. The C-NPU nomenclature used in Laboratory Medicine uses names and codes for the "system" in the patient, meaning the sample's origin – plasma, serum, urine, etc. However, the "system" is possibly inappropriate for reference materials intended for measurands in vitro. An appropriate naming and coding system for the matrix of secondary reference materials is needed, primarily if a modified sample matrix is used.

Proper ontologies are crucial for FAIR chemical data

For the present purpose, ontology is a formal representation of knowledge in the Chem/Bio area that can be shared among humans and computers. The current task of creating unique interoperable identifiers in the Chem/Bio area risks being underestimated as using InChI to describe the measurands and the "system" to illustrate the matrix. Advanced ontological considerations must be made from the outset to foresee and accommodate present and future demands for evolving and sufficiently advanced naming and coding of references in the Chem/Bio area for humans and computers (9-13). Amongst the future knowledge areas likely in need of addressing are identifiers for expressing the commutability of secondary reference materials, identifiers for expressing traceability, identifiers for characterizing processed sample matrices or measurands, and identifiers for expressing post-translational processing of measurands.

Therefore, knowledge and skills in the use of web ontology languages such as OWL 2 (https://www.w3.org/TR/owl2-overview/) are needed from the outset, both when applying already established naming and coding systems such as InChI and when considering designing new systems, e.g., for matrices (14, 15).

Recommendations from Day 1

1. To use InChI unique identifiers for measurands in primary reference materials
2. To use InChI unique identifiers for measurands in secondary reference materials combined with newly developed identifiers for sample matrix. The "system" names and codes from the c-NPU naming and coding systems may be initially considered.
3. To form a CCQM KCWG Task Group for determining how best to move forward with identifying and recommending ontologies/identifiers for the matrix of secondary reference materials that could be used in the KCDB and JCTLM databases, with

representatives from the CCQM KCWG, JCTLM DB WG, and external experts on web-ontology languages.

4. Later tasks for a CCQM KCWG Task Group could be chosen to create -
    a. identifiers for expressing the commutability of secondary reference materials,
    b. identifiers for expressing traceability,
    c. identifiers for processed samples/matrix, including, e.g. sample extracts and amplified genetic materials,
    d. identifiers for expressing post-translational processing (e.g. glycosylation) of large and complex molecules.
    e. databases for translating between InChI and other naming and coding systems such as Snomed-CT, LOINC and C-NPU.

References

1. Dybkaer R. Metrology in laboratory medicine - Reference measurement systems [English]. Accred Qual Assur 2001 Jan;6 1:16-9 as doi: Doi 10.1007/Pl00010430.
2. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI - the worldwide chemical structure identifier standard [English]. J Cheminformatics 2013 Jan 24;5 as doi: Artn 7 10.1186/1758-2946-5-7.
3. Rossander A, Lindskold L, Ranerup A, Karlsson D. A State-of-the Art Review of SNOMED CT Terminology Binding and Recommendations for Practice and Research [English]. Method Inform Med 2021 Dec;60:E76-E88 as doi: 10.1055/s-0041-1735167.
4. LOINC. Logical Observation Identifiers Names and Codes (LOINC) database accessible at http://www.regenstrief.org/ .). (Accessed.
5. Olesen H. Properties and Units in the Clinical Laboratory Sciences .1. Syntax and Semantic Rules - (Iupac-Ifcc Recommendations 1995) [English]. Pure Appl Chem 1995 Aug-Sep;67 8-9:1563-74 as doi: DOI 10.1351/pac199567081563.
6. Charlet P, Marschal A. Improvement in the traceability of environmental analysis by the relevant use of certified pure solutions and a matrix certified reference material [English]. Trac-Trends in Analytical Chemistry 2004 Mar;23 3:178-84 as doi: 10.1016/S0165-9936(04)00308-5.
7. Delatour V, Lalere B, Saint-Albin K, Peignaux M, Hattchouel JM, Dumont G, De Graeve J, et al. Continuous improvement of medical test reliability using reference methods and matrix-corrected target values in proficiency testing schemes: application to glucose assay. Clinica chimica acta; international journal of clinical chemistry 2012 Nov 20;413 23-24:1872-8. Epub 20120731 as doi: 10.1016/j.cca.2012.07.016.
8. Myers GL, Waymack PP. Matrix Effects in Biological Reference Materials Used in the Standardization of Cholesterol Measurements [English]. Fresen J Anal Chem 1990;338 4:538-42 as doi: Doi 10.1007/Bf00322533.
9. Arp R, Smith B, Spear AD. Building ontologies with Basic Formal Ontology. Cambridge, Massachusetts: Massachusetts Institute of Technology; 2015.
10. Dybkaer R. An Ontology on Property for Physical, Chemical, and Biological Systems, http://ontology.iupac.org/ontology.pdf. Copenhagen: Blackwell Munksgaard; 2004.
11. Dybkaer R, Storring PL. Application of IUPAC-IFCC recommendations on quantities and units to WHO biological reference materials for diagnostic use: recommendations 1994 [eng]. J Int Fed Clin Chem 1994 Jun;6 3:101-3. Epub 1994/05/08.

12. Dybkaer R, Solberg HE. Approved Recommendation (1987) on the Theory of Reference Values. Part 6. Presentation of Observed Values Related to Reference Values [English]. J Clin Chem Clin Bio 1987 Sep;25 9:657-62.
13. Dybkaer R. Vocabulary for use in measurement procedures and description of reference materials in laboratory medicine. Eur J Clin Chem Clin Biochem 1997;35 2:141-73.
14. Vita R, Zheng J, Jackson R, Dooley D, Overton JA, Miller MA, Berrios DC, et al. Standardization of assay representation in the Ontology for Biomedical Investigations [English]. Database-Oxford 2021 Jul 7 as doi: ARTN baab040 10.1093/database/baab040.
15. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, et al. The Ontology for Biomedical Investigations [English]. PloS one 2016 Apr 29;11 4 as doi: ARTN e0154556 10.1371 journal.pone.0154556.

Day 2 discussions: Digitalization of CRM Certificates

Issues which the workshop wished to focus on in Day 2

1. Gain understanding of the various options for producing and maintaining digital certificates and specific issues that need to be addressed for CRM certificates.
2. Gain understanding on how FAIR principles can be included in the production of digital CRM certificates. Understand the potential challenges within the scope of CCQM.
3. Gain understanding of client expectations on uses, benefits and challenges of working with digital CRM certificates.
4. Gain understanding on the level of resources and infrastructure required to develop and maintain digital certificates.
5. Develop recommendations on options for NMIs wishing to implement Digital CRM certificates and adhere to FAIR principles.

Main points of discussion

1. Use of standard vocabularies for both analyte and matrix. There are numerous ways of doing this. There are standardized approaches for analyte names – only some communities (e.g. IFCC-IUPAC) have tackled a common vocabulary for the matrix/material
2. Resources required to digitalise CRM certificates – level of effort varies depending whether it is one value reported or whether values are reported for different pressures and temperatures, for example.
3. Resources that a user needs to work with Digital RMCs: a generic tool should be developed to read certificates both for machines and humans. This should be harmonized so the tool can be used for certificates from all different suppliers. This may also extend to the provision of APIs and even code so that users can retrieve and analyse the data associated with the certificate.
4. Digitalization can go beyond the certificate itself and capture the information and data provided via the measurement methods as well as data processing steps. There are proposals to demonstrate this approach. It would be valuable to understand for which sectors or application areas this approach would be of most value, as the resource to achieve this is significant.
5. Security issue for DRMCs were raised including the ability to protect the data and make sure it has not been changed. Various levels of security were discussed from electronic signatures all the way to using encryption and Blockchain (for some specific potentially high risk applications) and keeping data secure from the RM all the way to the user.
6. User requirements and expectations for DRMCs were discussed. This is an area where it would beneficial for NMIs to approach their customers, learn and report back to the CCQM, in areas as diverse as Gas Standards for Industry, RMs for regulated industries such as Diagnostics or the newly developed Cannabis Industry to Biological RMs which are often also used for academics and discovery and for new measurements beyond their original

scope. Commercial producers of RMs see a future focus for analytical laboratories on automation, and associated time savings, and machine to machine readability , with digitalization of RM certificates feeding into this, with reduced time and reduced risk of manual errors. Distributors of different reference materials could be faced with the challenge of being able to provide APIs that would allow users to interrogate all the CRMs and associated data for the RMs they were distributing.

7. Additional challenges for automation include overcoming issues on lack of uniform data formatting, including from different instrument vendors. Working with various stakeholders toward a standardized data format would be highly beneficial to future evolution sin analytical processes.
8. Sustainability of efforts were discussed including data formats, software versions that change and may no longer be accessible, as well as those linked to RMs that maybe available for several decades. Larger (or national) repositories for data related to RMs may not be accessible to small producers of RMs, and as a result the question was asked where this data could be safely held? The sentiment was expressed that digitalization should be done when there was clear benefit of doing so, and that the resources spent should be proportional to the expected benefits, noting that in the long term the potential was for reduced costs, but an upfront investment was needed.
9. The auditability of the DRCMs was raised – and this is an essential requirement, especially for accredited laboratories.

Recommendations from Day 2

1) A group should be established to address harmonization of approaches to the production of DRMCs. The group should include relevant external stakeholders in addition to NMIs. The workshop recommended that this be a Task Group within CCQM and would also liaise with the CIPM FORUM-MD, noting that the production of DRMCs has sufficient difference with DCCs that it warrants an expert group.

2) Client expectations with regard to DRMCs need to be clarified as they are key for future developments including the level of accessibility to supporting data they would need. A plan for this needs to be identified, including engagement with stakeholders. This should be included in the terms of reference of a new CCQM TG, noting that NMIs will also be encouraged to report back on their assessments of stakeholder needs in various measurement communities.

3) Security issues in relation to DRMCs are an area of concern. Recommendations on how these can be addressed should be developed, and this will also be included as an activity within the terms of reference of the CCQM TG. Additional issues would be how to deal with modifications of certificates and their removal following the requirements of ISO standards.

4) Open and standardized data formats will be a key issue in digitalization efforts and will require dialogue and agreements with instrument vendors. Addressing this issue is a major undertaking. The CCQM TG should consider whether it can effectively start to address this, or develop a plan on how this could be done.

5) The resources to initiate and maintain a RM certificate digitalization effort need to recognized, and a roadmap developed on how the production of DRMCs will become accessible to NMIs that have resources to implement but not develop their individual solutions. The action will be included in the terms of reference of the CCQM TG.

Day 3 discussions: Application of FAIR principles to reference data and databases

Issues which the workshop wished to focus on in Day 3:

1. Gain understanding of the benefits and challenges to transitioning to the use databases for scientific data.
2. Gain understanding of applying FAIR principles to Database development. What are the benefits and challenges?
3. Gain understanding on examples of how FAIR principles have been applied to various databases, libraries, or any other data products.
4. Develop recommendations on key issues to be considered in database development and maintenance when wishing to apply FAIR principles.

Overview of discussions:

The pre-recorded presentations presented for this session included a very good overview of what FAIR principles are when it comes to the management of data and what needs to be considered to apply it to databases and the management of reference data by Bob Hanisch from NIST. The rest of the presentations consisted of examples of different databases in the field of chemical measurement and how the FAIR principles were applied in these databases.

These examples include a presentation by Adriaan van der Veen from the Van Swinden Laboratorium (VSL) in the Netherlands on the development over the years of the database being used to manage their large suite of primary reference gas mixtures (PRGMs). The presentation also pointed out the lack of adherence of some of the FAIR principles such as the use of unique identifiers. Evan Bolton gave a presentation on PubChem where extensive progress has been made on the development of unique identifiers and the application of most of the FAIR principles with the caveat that it applies to a only a subset of the components and matrices that apply to metrology in chemistry and biology.

Other examples included Chemotion that is used for Electronic Lab Notebooks where the emphasis is place on the findability and accessibility of data to be able to reuse it. Interoperability of data was highlighted as a difficult aspect of the FAIR principles to apply. Carsten Kettner gave an overview of the Strenda DB that was developed for enzymology data which makes use of other databases such as PubChem. The presentation by Ben Place on the DIMSpec Project for mass spectrometry data highlighted once again the issue of interoperability of data as it applies to the issue of the protection of the intellectual property of instrument vendors as it applies to their proprietary instrument software and the need for instrument vendors to be willing to adopt open-source formats for data transfer from analytical instruments. The presentation by Stuart Clark on the IUPAC Gold Book also highlighted the challenges of finding unique identifiers for complex systems when the field of application is wide such as the measurements related to chemistry and biology.

Outcomes and recommendations:

The session on Day 3 of the workshop provided a good overview of what the FAIR principles are and how they are applicable to the good management of databases/repositories and other forms of reference data. The collection of examples of data repositories also explained very practically how the FAIR principles can

be applied to data management and already highlighted some of the limitations and issues for discussion during the workshop.

The issue of the need for unique identifiers were recognised very early in the presentations and were discussed quite extensively throughout the workshop. The need for harmonised terms, nomenclature and ontologies for chemical and biological measurements were one of the first recommendations. As already discussed during Day 1 of the workshop the combination of components in a complex matrix as it applies to chemical and biological measurements adds an additional level of complexity to the issue of unique identifiers for chemistry and biology.

The recommendations form the discussions during the workshop on Day 3 include the following:

- Formation of a small group of experts to investigate the application of FAIR principles to the KCDB and JCTLM databases
- Collection of a list of all (if possible) ontologies and nomenclatures applicable to chemistry and biology CMCs
- Preparation of a summary document with information and a collection of resources (websites, publications, etc.) for NMIs/DIs interested in developing FAIR data repositories
- Investigate the possibility of sharing repository resources withing the NMI community
- Consider the development of a database for all CCQM comparison results beyond was available with the current version of the KCDB