Analytica Chimica Acta 642 (2009) 3–5

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

IUPAC project: A glossary of concepts and terms in chemometrics

D. Brynn Hibbert[a,▫], Pentti Minkkinen[b], N.M. Faber[c], Barry M. Wise[d]

[a] School of Chemistry, University of New South Wales, Sydney, NSW 2052, Australia
[b] Lappeenranta University of Technology, Department of Chemical Technology, P.O. Box 20, FIN-53851 Lappeenranta, Finland
[c] Chemometry Consultancy, Rubensstraat 7, 6717 VD Ede, The Netherlands
[d] President, Eigenvector Research, Inc., 3905 West Eaglerock Drive, Wenatchee, WA 98801, USA

ARTICLE INFO

Article history:
Received 7 November 2008
Received in revised form 14 February 2009
Accepted 17 February 2009
Available online 21 February 2009

Keywords:
IUPAC
Chemometrics definitions
Chemometrics terms
Chemometrics concepts
Wiki

ABSTRACT

A project has been initiated by the International Union of Pure and Applied Chemistry (IUPAC) to create a glossary of concepts and terms in chemometrics. This will be accomplished by consultation with the community through the means of a wiki – a web site that can be modified by users (see http://www.iupacterms.eigenvector.com/index.php?title=Main_Page). Over time new terms can be added, and consensus definitions arrived at. The definitions will be published as IUPAC recommendations.

1. Introduction

Chemometrics is defined by the International Chemometrics Society as "the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods", but many other variants may be found in the literature highlighting aspects of analytical chemistry, statistics, information, and data analysis. It has somehow managed to maintain its status along with, but somehow independent of, established fields of chemistry and even analytical chemistry. The sometimes lack of understanding between chemometrics and metrology communities was encapsulated at a recent editorial board meeting of Accreditation and Quality Assurance in Chemistry, when a member was minuted as opining, in response to suggestions for widening the scope of the journal, that "... the term 'chemometrics' might put off many people".

The tight focus of the subject has served chemometrics very well. With a core of scientists who have devoted their careers to advancing different methods, chemometrics has a clear footprint [1,2], albeit with a terminology that has grown through different areas of application. The small size of the community has led to a situation in which assiduous reading of the literature allows a reasonable understanding of what is meant, even when there is a conflicting terminology. However with growth into many fields, and a maturing of the subject it may be now timely to tease out the fundamental concepts of the subject and define appropriate terms. This is certainly not a new aspiration. In 1983, Bruce Kowalski wrote in a preface to the proceedings of the Consenza conference. "A primary problem that surfaced constantly was the lack of a commonly agreed notation and nomenclature for chemometrics" [3]. It was suggested at the time that a glossary of terms be made (a "slate of chemometrics notation") and submitted to IUPAC. This appears not to have happened.

The present IUPAC project will attempt to bring together terminology as used by practicing chemometricians and offer a consistent vocabulary[4]. As one of the members of the task group (BMW) writes "authors should be free to use terms as they wish provided that they define them unambiguously in their text", and who can gainsay this? However a series of papers with authors unambiguously defining terms that are quite different from those found in other papers, can lead to confusion. Add to this the propensity for instrument manufacturers who bundle software with their products to re-invent concepts with their own proprietary terms and it might be said that over the years the field of chemometrics has developed a workable, but perhaps sub-optimal, vocabulary.

2. Chemometrics

As this subject has been of great interest to the chemometrics community and is rehearsed every now and again, only a brief

▫ Corresponding author. Tel.: +61 2 9385 4713; fax: +61 2 9385 6141.
E-mail addresses: b.hibbert@unsw.edu.au (D.B. Hibbert), pentti.minkkinen@lut.fi (P. Minkkinen), nmf@chemometry.com (N.M. Faber), bmw@eigenvector.com (B.M. Wise).

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.02.020

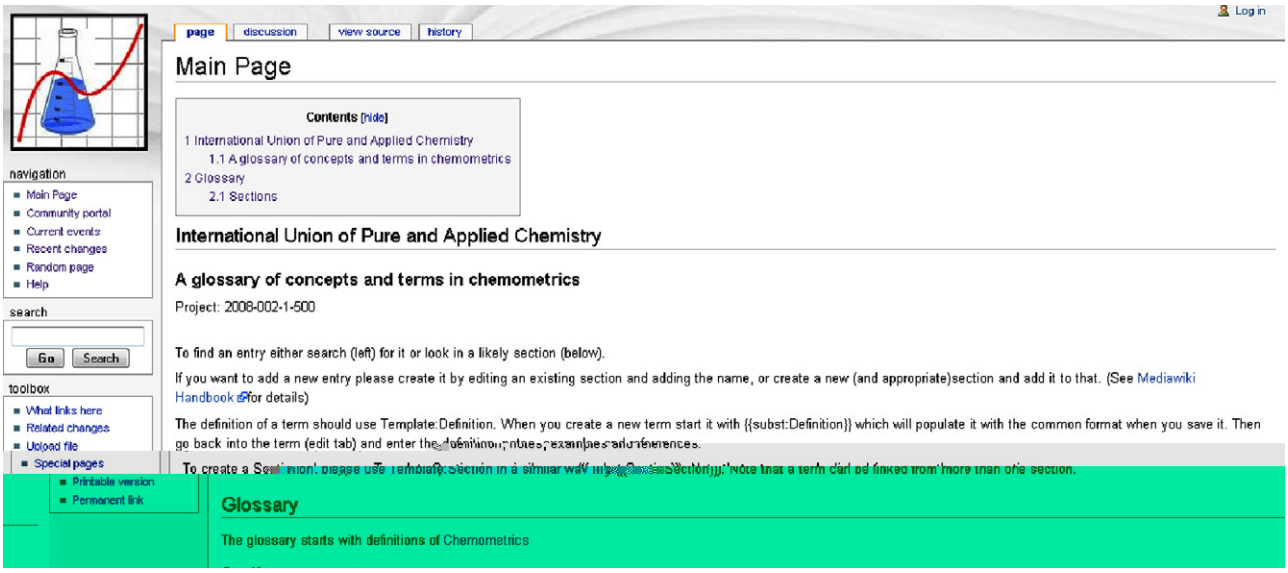

Fig. 1. The IUPAC project wiki page on 'chemometrics' (http://www.iupacterms.eigenvector.com/index.php?title=Main_Page), accessed 14 February 2009.

resume of the early years will be given here [5–7]. Despite the observation that any use of statistics, or even mathematics, allows every chemist from Lavoisier to Box via Gossett, Youden and Fischer to be called chemometrician, the start of chemometrics is held to be in the early 1970's with Wold's coining of the term 'Kemometri' (see [8] for a history of the words used for chemometrics), although papers by Kowalski, Jurs and Isenhour in Analytical Chemistry in 1969 set the scene [9]. The division between univariate and multivariate data seems to drive the new science of chemometrics, and breakthrough work in pattern recognition, multivariate calibration and optimization cemented its position. The establishment of the International Chemometrics Society, two international journals, long running conferences (CAC, SSC, ICRM, MULDAST, etc.), and various activities, has allowed chemometrics to extend in many directions. There have been some attempts to create vocabularies, but they have tended to be used in restricted circles, and have not been universally agreed. Possibly the most comprehensive is the web-based *Chemometricopendium* [10] developed by Vandeginste, and which will form the starting point of many of the entries in the proposed wiki.

3. IUPAC

The International Union of Pure and Applied Chemistry (IUPAC) was formed in 1919 by chemists from industry and academia, in response to concerns about the lack of consistent terminology and nomenclature. Since then, IUPAC has long been recognized as the world authority on chemical nomenclature, terminology, standardised methods for measurement, atomic weights and many other critically evaluated data. With membership of national science organizations from 77 countries IUPAC claims to be truly representative of world-wide chemistry. In recent years IUPAC has operated through the project system in order to fulfil its Strategic Plan. Projects must also satisfy at least one of the following key criteria:

1. They should be related to the needs of the chemists in the world, not just in a country or a region.
2. They should be related to the role of chemistry for the needs of mankind.
3. They should best be approached by an international team such as IUPAC can assemble.

In late 2007, a project proposal "a glossary of concepts and terms in chemometrics" was submitted in the names of the authors of this paper. The project was approved in early 2008, and the first meeting was held at the CAC2008 in Montpelier, France in July.

4. The Wiki approach

For a subject with diverse users of chemometric concepts a working vocabulary is not easy to arrive at. There would be no point in any group, no matter how august, writing down a definitive set of terms, because there would always be some who would assert their rights to use their existing terms, or even produce yet more neologisms. However if all definitions of terms can be subjected to worldwide scrutiny, then there is a chance that a consensus can emerge.

The ubiquitous 'wiki' is an ideal tool for this public airing of terms in a vocabulary. It has been successfully used in an IUPAC project by Kermit Murray, who led a project "Standard Definitions of Terms Relating to Mass Spectrometry" [11]. The format was very popular with 5000 hits per day and a useful vocabulary being produced.

A wiki website works best for content that consists of short entries on well-defined subjects, such as found in dictionaries, encyclopedias and glossaries. A page contains one entry (definition) with links to other pages and external references. Meta data is easily added for grouping terms by application, mathematical approach, and inclusion in software and so on. A history of edits is maintained to allow understanding of the evolution of the definition.

The archetype, Wikipedia, is known for its entirely open access, with attendant concerns of accuracy, and manipulation for ulterior motives. However it is possible to restrict the ability to edit to registered users, or to implement an editorial or peer review of entries.

The IUPAC wiki is hosted by Eigenvector[1] at http://www.iupacterms.eigenvector.com/index.php?title=Main_Page. The first entry is, appropriately, a definition of chemometrics (Fig. 1).

5. Writing terms and definitions

Definitions of terms are written according to rules that ensure a consistency of approach and format [12]. For example the definition

[1] Eigenvector Inc., have kindly hosted the IUPAC wiki. However the company asserts no editorial control over the wiki and neither endorses, nor derives particular benefit from the wiki.

of a term should be able to be substituted for the term in use. This means that definitions do not start with an article. Definitions that contain terms that themselves are defined, should be able to have the definitions expanded within them, that is all definitions must be recursive. Definitions should not be circular, and notes and examples can be added, but should not form part of the definition. Vocabularies should respect higher order compilations, for example the international vocabulary of terms in metrology (VIM) [13], or statistics–vocabulary and symbols [14].

It will be the responsibility of the task group to eventually bring together the terms in the wiki and write a set of recommendations to be published in Pure and Applied Chemistry.

6. The IUPAC project

The project team will establish a wiki site that initially will be populated with terms from existing lists. How these will be grouped, by method (spectroscopy, electrochemical) or application (environmental, forensic, process control) is to be determined. Also at what level of abstraction – 'chemometrics' is the highest, 'multivariate analysis' is quite high, but 'PCA' and 'scores' will appear at a lower level.

Terms will be requested and proposed. Definitions will be solicited, and multiple definitions allowed. It is important to link definitions or usages of terms with applications and areas of chemometric practice. After several months of operation the task group will critically review the site and propose a definitive vocabulary. It is likely that the scope of the project will widen and further projects will be proposed. It is the intention of the task group that a vocabulary will not be proposed as an IUPAC recommendation until there is sufficient consensus within the chemometrics community.

7. Conclusion

The IUPAC task group calls upon the chemometrics community to engage in this process that will result in a vocabulary of terms that can be used, knowing that there is a consistent set of chemometrics terms that can provide a basis for future research. We ask

1. Please consult the wiki and comment on terms that are there.
2. Contribute new terms that you believe should be added.
3. Give feedback to the task group, and publicise the project.

Acknowledgments

The task group acknowledges the kind help of Bernard Vandeginste for making his chemometricopendium (http://www.vicim.com/chemometrics%20thesaurus.web/index.html) available to populate the wiki, and Eigenvector Inc. for hosting the wiki.

References

[1] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S.D. Jong, P.J. Lewi, J. Smeyers-Verbeke, Data Handling in Science and Technology, Elsevier Science B.V., Amsterdam, 1997, 207.
[2] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S.D. Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier Science B.V., Amsterdam, 1998.
[3] B.R. Kowalski, Chemometrics mathematics and statistics in chemistry, in: B.R. Kowalski (Ed.), NATO Advanced Study Institute on Chemometrics Mathematics and Statistics in Chemistry, Springer, Cosenza, Italy, 1983, 485.
[4] D.B. Hibbert, P. Minkkinen, B.M. Wise, A glossary of concepts and terms in chemometrics, International Union of Pure and Applied Chemistry (2008).
[5] K. Esbensen, P. Geladi, Journal of Chemometrics 4 (1990) 389.
[6] P. Geladi, K. Esbensen, Journal of Chemometrics 4 (1990) 337.
[7] S.D. Brown, Chemometrics and Intelligent Laboratory Systems 30 (1995) 49.
[8] R. Kiralj, M.M.C. Ferreira, Journal of Chemometrics 20 (2006) 247.
[9] B.R. Kowalski, P.C. Jurs, T.L. Isenhour, C.N. Reilley, Analytical Chemistry 41 (1969) 695.
[10] B.G.M. Vandeginste, Chemometricopendium, a chemometrics thesaurus, Virtual Institute of Chemometrics and Industrial Metrology (VICIM) (2004).
[11] K. Murray, Web Glossaries the Wiki Way, Chem. Int., in: IUPAC, 2007, p. 23.
[12] ISO, Terminology work – vocabulary – part 1: theory and application, in: ISO (Ed.), 1087-1, International Organization for Standardization, Geneva, 2000.
[13] ISO/IEC, International vocabulary of metrology – Basic and general concepts and associated terms VIM, Guide 99 (in the name of BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML), International Organization for Standardization, Geneva, 2007.
[14] JCGM, International vocabulary of metrology – Basic and general concepts and associated terms VIM, JCGM 200 (in the name of BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML), BIPM, Sevres, http://www.bipm.org, 2008.