

## ESSAY

# The CODATA Role in Promotion of Data Quality

Remarks on Receipt of the 2016 CODATA Prize

David R. Lide

Retired, US  
drlide@post.harvard.edu

Promotion of data quality was a key feature of CODATA's original charter and should continue to receive the highest priority.

**Keywords:** CODATA; founders'; data; quality; evaluation

It is a special honor to receive the CODATA Prize on the 50<sup>th</sup> anniversary of the birth of CODATA. My association began when I attended the First International CODATA Conference in Arnoldshain, Germany in 1968, just two years after CODATA's founding. That was a remarkable event, probably the first international conference to deal with scientific data issues in a generic sense, and we owe a debt of gratitude to CODATA's founders. They were all distinguished scientists, highly respected for their own research contributions and service to their governments, and they had the foresight to recognize that science was at the beginning of a major change, where the rapidly increasing production of new data from experiments and observations threatened to swamp the traditional archival publication mechanisms, leading to the possibility that important data would be lost to future generations of scientists.

The six founding fathers of CODATA were:

- Frederick Rossini (United States) – a physical chemist who had developed techniques for highly accurate measurement of the heat changes associated with chemical reactions. He had also established programs for evaluating thermochemical data reported in the literature and selecting the most reliable values.
- Boris Vodar (France) – a Russian émigré to France whose specialty was the measurement of properties of materials at very high pressures.
- Sir Gordon Sutherland (UK) – a Scotsman who did early work on infrared spectroscopy and the application of quantum mechanics to the determination of the three-dimensional structure of molecules.
- Wilhelm Klemm (Germany) – a distinguished chemist who made many contributions to inorganic chemistry and took part in early efforts for systematic compilation of chemical data.
- Masao Kotani (Japan) – who began his career in the early days of modern atomic physics and later did pioneering work in biophysics.
- Mikhail Styrikovich (Soviet Union) – a physicist and engineer who specialized in fluid dynamics and heat transfer, and who was keenly aware of the need for accurate data in the design of industrial processes.

Although these leaders had worked in different scientific fields, they all understood the prime importance of accuracy and reliability in the data they dealt with, whether that data was to be used in basic scientific research or for industrial applications. Thus data quality became a core objective as CODATA began to develop its programs.

I want to make several comments about these founders, and about the first CODATA Conference itself:

1. They were all physicists and chemists, and virtually all the attendees at the first Conference were from the physical sciences. This would change as CODATA evolved to cover the other scientific disciplines that ICSU included in its charter.
2. They all came from the industrialized countries of Europe, North America, and Japan, as did every one of the 90 conference attendees. There were no attendees from developing countries. That also would change.
3. None of the founders had any experience with computers. In fact, some were a little nervous that computer specialists would gain too much influence in CODATA. Nevertheless, there was one session at the first Conference on the use of computers for data management at which Olga Kennard, who founded the Cambridge Crystallographic Data Centre, was an active participant.
4. The Conference included many sessions on methods for critical evaluation of data in different fields. In some, such as crystallography and nuclear physics, methodology for assessing data quality was already in use. Efforts of this type were just beginning in other fields. The Conference served the very useful purpose of bringing together data specialists in different disciplines who otherwise had no contact with each other. The emphasis on data quality also influenced the first projects CODATA set up, such as the Task Group on Fundamental Physical Constants.

Thus my association with CODATA began, and was to continue for the next 25 years. During those years, I witnessed its evolution into an influential scientific organization and saw the expansion of its universe. In terms of disciplinary scope, CODATA extended first into the biosciences, when we started projects like the Protein Sequence Data Task Group and the Hybridoma Data Bank. Several small projects were started in the geosciences, and more came later. And increasing attention was focused on data of industrial interest, such as properties of engineering materials. Attention to data quality was a feature of all these new activities.

CODATA also expanded in geographical terms, from 7 National Members in 1968 to 22 members at the 1990 Conference – and even more today. I was especially pleased that the Chinese Academies at Beijing and Taipei joined CODATA during my tenure as President. And it is gratifying to see that so many countries from Asia and Africa are now participating in CODATA – it is quite a change from 1968.

And, of course, during these years computer technology has become an integral part of virtually every CODATA activity. I'd like to note that CODATA produced its first computer-based product, called the CODATA Referral Database, in 1990. This was a guide to data sources in different disciplines, distributed on floppy disks with integrated search software for use on a personal computer – and this was before the Internet and Google were invented. I am proud that my wife Bettijoyce, who is here today, was a leader in producing that first CODATA electronic database.

During my years with CODATA I met and worked with hundreds of people who took part in the various CODATA activities. I have warm memories of these colleagues, who deserve the credit for the progress CODATA made in its first quarter-century. There are far too many to mention by name, but I do want to acknowledge one person, Phyllis Glaser, who served as Executive Director for 20 years. Phyllis gave her heart and soul to CODATA and was often the glue that held us together during turbulent times. And her esoteric flat on the Rue Bleue served as a stop-over for many children of CODATA participants as they back-packed through Europe. Those of us who worked with her were saddened when Phyllis passed away in late 2014.

In conclusion, CODATA began at a time when data evaluation and data management were considered rather dull subjects by most scientists. I was Director of the Standard Reference Data program at the National Bureau of Standards at that time, and I had a constant struggle with my management to get money to support data evaluation and to get proper respect for the people who were involved in data work. Fifty years later, it is a very different world. "Big Data" has become a popular buzz word, and many organizations want to get involved. "Data Scientist" is now a recognized profession. Many new organizations have been set up to deal with data issues. It is certainly pleasing to see the increased attention now given to the storage, preservation, and retrieval of scientific and technical data, and to removing the barriers to accessing this data. However, I worry that one thing is not getting proper attention, and that is the need for assurance of data quality. The Internet is a great tool for finding data, but if I Google for the thermal conductivity of magnesium at -50 Celsius, I might find 6 different values on 6 different sites. To justify the investment in massive data archives, it is essential to build into these structures a process for evaluating the quality of the data that

goes in – to include a method for selecting the most accurate data and documenting the conditions under which the data were obtained. Data quality was a core objective when CODATA was founded, and is still the first objective in the CODATA charter. I want to conclude my remarks with a plea to everyone involved with data today – please continue the CODATA tradition of recognizing data quality as the highest priority.


### Competing Interests

The author has no competing interests to declare.

**How to cite this article:** Lide, D R 2018 The CODATA Role in Promotion of Data Quality. *Data Science Journal*, 17: 3, pp. 1–3, DOI: <https://doi.org/10.5334/dsj-2018-003>

**Submitted:** 18 October 2016    **Accepted:** 04 December 2017    **Published:** 24 January 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 