

RESEARCH PAPER

Building an International Consensus on Multi-Disciplinary Metadata Standards: A CODATA Case History in Nanotechnology

John Rumble¹, John Broome² and Simon Hodson³¹ R&R Data Services, Gaithersburg, MD, US² John Broome, Ottawa, CA³ CODATA, Paris, FRCorresponding author: John Rumble, Jr. (john.rumble@randrdata.com)

Science today is rapidly becoming both multi-disciplinary and data-driven. These two trends pose new challenges to the capture, management, sharing, and dissemination of research data. Multi-disciplinary science means diverse data generation communities and equally diverse user groups. Data-driven means that sharing data among different communities is more important than ever because of the growth of modeling and knowledge discovery. Nanotechnology is a prime example, involving chemistry, physics, materials science, toxicology, environmental science, and many other disciplines. During the past few years, CODATA has created an international, multi-disciplinary Working Group that has developed a number of critically important metadata standards to facilitate sharing nanomaterials data. In this paper, we discuss the challenges faced in starting and executing this work, as well as the approaches taken to make progress on producing internationally accepted metadata standards. Many of these approaches are directly applicable to other multi-disciplinary subjects.

Keywords: Metadata standards; data standards; international scientific cooperation; scientific data; nanomaterials; nanoinformatics

1. Introduction

As science becomes more multi-disciplinary and data-driven, new challenges arise with respect to the capture, management, sharing, and dissemination of research data. Multi-disciplinary science means diverse data generation communities and equally diverse user groups. Data-driven means that sharing data among different communities is more important than ever because of the growth of modeling and knowledge discovery.

In this paper, we will explore these challenges and describe how The Committee on Data for Science and Technology (CODATA) has approached developing usable solutions for defining metadata standard for multi-disciplinary data. We discuss the efforts of the CODATA “Working Group on Nanomaterials” (WG) and its successful development of internationally-accepted metadata standards. We also abstract from this work an approach that is usable for similar multi-disciplinary efforts, as well as identify challenges that must be overcome regardless of the subject.

The structure of this paper is as follows:

Section 1: Introduction

Section 2: General problems of multi-disciplinary metadata standards

Section 3: International dimensions of metadata

Section 4: Specific problems associated with nanomaterials metadata

Section 5: The approach used by the CODATA WG

Section 6: Challenges

Section 7: Applicability to Other Disciplines

Section 8: Conclusions

2. General problems of multi-disciplinary metadata standards

The term *metadata* is usually defined as data about data. In the context of data on the properties of substances and objects of interest to science and technology (S&T), metadata are the piece of data and information that allows a user to understand the generation, meaning, and quality of S&T data. Examples of important types of metadata in S&T include:

- Units
- Uncertainty
- Method of measurement
- Equipment or instrument used
- Measurement method
- Measurement conditions
- Data analysis
- Etc.

A number of problems arise when developing metadata standards for multi-disciplinary data, some of which are listed below. Not all are relevant in every multi-disciplinary situation, and others, not listed, may also apply. What is important is the recognition that developing metadata standards must often overcome significant challenges. The examples given are drawn primarily from the physical, chemical, materials, and engineering sciences, but occur in every scientific and technical discipline.

Terminology inconsistency: Different disciplines develop their internal terminology irrespective of other disciplines, and trying to harmonize terminology across disciplines requires careful consideration of the equivalency of terms and which term should be chosen among different possibilities. An example is the harmonization of terminologies for describing a chemical substance developed for two distinct disciplines – genomics and chemistry (Hill et al. 2013).

Scientific viewpoint: Different properties of the same substance may be of importance to different disciplines, with the result that metadata types have different degrees of depth. For example, in nanotechnology, toxicologists are interested in the solubility and toxicity of a nanomaterial, while materials scientists are interested in the physical and electronic properties of the same nanomaterial. The metadata associated with those properties are quite different, and different sub-disciplines in nanotechnology have developed seemingly incompatible metadata (Boverhof et al. 2015).

Language: Even though English is recognized as the common language of science, metadata standards often must deal with other languages (Sayle, 2009).

Formality of existing standards: when multi-disciplinary data work begins, there are usually existing metadata standards developed by one or more of the involved disciplines; their formality may range from adopted national and international standards to informal working conventions. A successful effort that consolidated formal and informal standards in various astronomy disciplines is the Virtual Observatory, which began as a vision (Quinn et al. 2004) and has since become a critical tool for all astronomers (Hanisch, 2015).

Lack of scientific knowledge: In many instances, the independent variables governing a specific measurement result or property may not be well known. This is especially common in emerging scientific fields or when a new property is determined to be interesting or relevant (Marchese-Robinson et al. 2016).

Complexity of metadata: Some disciplines may be further along in identifying and defining the independent variables and other metadata associated with a property or other data type. The development of metadata standards in crystallography has taken place over decades, and the level of technical metadata covered by those standards is extremely detailed and precise (Hall and McMahon, 2005).

Type of research organizations: Some disciplines are large science, with research conducted in teams at large facilities; others are small science with many individual researchers performing investigator-initiated, peer-reviewed research in small laboratories with small groups with little impetus for developing standards. A recent paper examined this situation in chemistry (Bruno and Frey, 2017).

Source of funding: Some disciplines have funding sources concentrated in a few centralized agencies, usually at the national level, that can encourage or demand standardization; other disciplines feature a wide variety of funding sources without such pressures. One can compare, for instance, the situation in astronomy, with large funders driving standards development (Quinn et al. 2004) (Hanisch, 2015) with that of chemistry, which features research in small groups funded under diverse programs (Bruno and Frey, 2017). This factor does not uniformly affect “small” science, as crystallography, which features funding diversity as great as that found in chemistry, does have well defined metadata standards (Hall and McMahon, 2005).

Balance between academia and industry: Academia usually does not see a business case for standardization, unlike industry where standards are common; instrument manufacturers are comfortable with using standards to capture data, unlike many academic researchers who are concentrating on knowledge discovery. Data on engineering materials is a case in point (Rumble, 2017).

Wide variety of user communities: Ultimately scientific data are saved and made available to diverse user communities that have different research and technological goals, which are not always participating in the development of metadata standards; consequently their needs may not be addressed in the development process (Karcher et al. 2018).

3. International dimensions of metadata

A number of international issues related to metadata standards have been outlined in Section 2, but three issues require additional discussion. These issues are not necessarily barriers or problems, but help define the framework in which work of this type is done.

International scientific unions: Most scientific disciplines have a well-established international scientific union, many of which have some kind of activity related to data. In many cases, the unions have already developed some metadata standards relevant to the work being considered. These unions are a source of experts, expertise, and knowledge about their discipline specific data situations. Most of the unions are members of the International Science Council (ISC; formerly “International Council for Science” – ICSU) and ISC can help reach out to them in identifying contacts and active experts. At the same time, the unions are very much focused on discipline-specific data issues and may not have experience in working across disciplines.

Standards development organizations: Most countries have one or a small number of formal national standards development organizations (SDOs); the United States is the outlier with over 600 recognized national bodies (ANSI SDOs, n.d.). On the international and regional scale, SDOs, such as ISO (“ISO – International Organization for Standardization,” n.d.), IEC (IEC, n.d.), ASTM (ASTM International, n.d.), and CEN (CEN, n.d.), exist and play a major role in metadata standards. Most of these organizations support their activities by selling their standards documents, often for a considerable fee. Intellectual property rights (IPR) are of importance when working with such bodies.

Individual expertise participation: The culture and funding availability to support individual expert participation in metadata standards development varies tremendously from country to country and region to region. While the Web has greatly reduced travel costs, face-to-face meetings still are important as well as the time required for participation. Engaging the correct expertise – discipline data experts and users alike – is critical to success; practical issues related to individual participation can present considerable difficulties. This subject is raised again below when we discuss funding.

4. Specific problems related to nanomaterials metadata

In 2013, a series of interdisciplinary workshops (CODATA, 2018) were sponsored by the International Council for Science (ISC) (ICSU, 2018) and CODATA, the ISC Committee on Data for Science and Technology (CODATA, n.d.). As a result, the CODATA Working Group on Nanomaterials was established to address the need for a more precise system to describe nanomaterials. Nanomaterials are defined as substances with one or more dimensions in the range of one to 100 nanometers, meaning that they are larger than most individual molecules yet significant different from common bulk materials such as metals, alloys, ceramics, etc. because of quantum mechanical effects exhibited by their surfaces and large surface to volume ratios. The commercial, scientific, medical, and technological value of nanomaterials is readily evident (Roco, 2011), and much exciting research has been funded in the last two decades (Dale et al. 2015).

At the same time, because of their size and reactivity, nanomaterials are viewed as potentially having undesirable effects on living organisms, either individually or through concentration in the environment or various food chains (Dale et al. 2015a) (Dale et al. 2015b). Consequently, groups interested in nanomaterial property data include disciplines involved in their design, optimization, and manufacture – chemistry, physics, materials science, biochemistry, and pharmacology – as well as disciplines interested in possible harmful effects – toxicology, ecology, medicine, and environmental science.

We describe below how these diverse communities were brought together to develop some of the needed metadata standards. Here we discuss a specific metadata problem identified in the aforementioned workshops, that is, the lack of a robust method for describing individual nanomaterials and collections thereof. Practically, the rich tools of chemical nomenclature were sufficient to describe limited aspects of

nanomaterials – e.g., their chemical and crystallographic structure – but prominent features such as size, shape, surface conditions, interactions, and topology had not well accepted and systematic methodology for description.

The CODATA WG undertook the development of the **Uniform Description System** for materials on the Nanoscale (UDS). This work involved defining the major descriptive features needed to identify a nanomaterial uniquely. In addition, the description system was to support the determination of the equivalency of two nanomaterials so their property data sets could be combined in a scientifically meaningful manner.

The problem of not having an internationally agreed upon, multi-disciplinary system to describe objects for which property data are important is not unique to nanomaterials. In biology, environmental science, materials science, and other fields, this problem is common. Hopefully the approach taken by the CODATA WG is applicable to similar situations.

5. Approach used by the CODATA WG

For over 50 years, CODATA has brought together international experts to work on scientific and technical data issues. Consequently, when asked by ISC to work in the area of nanomaterials data, CODATA drew on its experience with similar data issues to define the problem(s) of interest and to assemble the expertise necessary to develop solutions. Below we describe in detail CODATA's approach, recognizing that the specific data issues concern nanomaterials, but that the approach is extendable to other multi-disciplinary data areas.

5.1. Scoping the problem

As mentioned above, the intense scientific interest in nanotechnology has fostered much research aimed at characterizing existing nanomaterials and exploring new ones. This research has generated large amounts of property data of interest to a number of disciplines, and one challenge was to make those data as useful as possible to the broad community. To define the major data issues more clearly, CODATA started with a three-fold effort that included workshops, survey, and discussions with experts. Working with ISC, CODATA identified 13 international scientific unions (hereafter simply referred to as unions) and convened two international workshops in Paris with approximately 50 attendees at each workshop. In addition, CODATA used the results of a 2008 workshop held in Oak Ridge TN on the topic of nanoinformatics. At these workshops, a core problem was identified, that there was no consistent, harmonized, and international accepted method for describing nanomaterials, which meant that sharing data across research communities and disciplines was difficult. To confirm this conclusion, CODATA used a written survey responded to by over 50 international experts and conducted face-to-face or telephone interviews with about another 20 experts.

5.2. Setting up a Working Group

Developing a nanomaterials description system, which is in fact a metadata problem, was a reasonable task, so in 2013, CODATA formally set up its Working Group on Nanomaterials. The first task was to recruit international experts on a multi-disciplinary scale. To do this CODATA approached the international scientific unions (see **Table 1**) who identified representatives to participate. In addition, the WG leaders, who had worked extensively in nanotechnology, drew on their own expert connections to recruit additional participants.

It should be noted that CODATA Task Groups and Working Groups have for decades been set up to address data-related issues such as this one. One major reason these groups have been so successful is that CODATA provides a neutral umbrella under which experts can work with minimal nationalistic and institutional concerns, such as getting credit. Further, CODATA members span the entire global, which makes it easier to bring together international perspectives and expertise.

5.3. Obtaining seed funding

The work done by the CODATA WG required two types of funding: travel funds to bring experts together for face-to-face meetings and funds to support the time of individuals who do the majority of work outside meetings. For the WG, initial seed funding for travel and meetings was provided by ISC, through a small grant, and CODATA itself. Support for individual experts was provided by their organizations or piggybacked on related existing grants. Support was provided by some international unions to help for travel by their representatives.

The salient point here is that the clear importance of the work catalyzed support from a number of interested parties and did not require obtaining a large central grant or contract, which takes time to propose and obtain. This is important to maintain the enthusiasm of experts once a data problem is identified.

Table 1: International Scientific Unions That Participated in the CODATA WG.

Union Acronym	Union Name
IUPAC	International Union of Pure and Applied Chemistry
IUCr	International Union of Crystallography
IUTOX	International Union of Toxicology
IUFoST	International Union of Food Science and Technology
IUNS	International Union of Nutritional Sciences
IUPHAR	International Union of Basic and Clinical Pharmacology
IUPESM	International Union for Physical and Engineering Sciences in Medicine
IUMS	International Union of Microbiological Societies
IUPAB	International Union for Pure and Applied Biophysics
IUCN	International Union for Conservation of Nature
IUTAM	International Union of Theoretical and Applied Mechanics
IUBS	International Union of Biological Sciences
IUBMB	International Union of Biochemistry and Molecular Biology
IUMRS	International Union of Materials Research Societies
IUPAP	International Union of Pure and Applied Physics
IMU	International Mathematical Union
IUPS	International Union of Physiological Science

5.4. Providing leadership

One of the critical factors in the success of a CODATA WG is the availability of appropriate leadership expertise and support for those persons. The skill set required includes project management, contract management, scientific and data knowledge, and international science experience. The first two skills are discussed in more detail below. Because a WG brings together many different scientific and data experts, the knowledge the leader must have must be sufficient to lead the process and to achieve a consensus when conflicting opinions arise. Having experience in international science is important because working groups include experts of different international backgrounds. The same is true for having a multi-disciplinary background.

In this project, WG leadership was provided by a combination of nanomaterials expertise and data and metadata experience. All three WG leaders also had considerable experience in international collaboration and previous contacts with similar efforts in other countries. In today's global science environment, such experience is common.

5.5. Doing the work

At the start of such work, setting clear goals is extremely important. In almost every situation for setting metadata standards, the breadth of the effort is almost endless. Without clear goals and specific targets, making progress can be difficult. Goals, however, cannot be made in a vacuum and can realistically be set only upon an initial survey of user needs and assessment of the complexity of metadata involved. While much of the work today can be accomplished by teleconferences and web-based meetings, the importance of face-to-face meetings is hard to over-emphasize. These need to be structured to allow for a free flow of ideas and discussions of options. In the case of the CODATA WG, meetings were held about once a year and in different regions (North America, Europe, and Asia) to reduce travel burdens. Face-to-face meetings were extremely useful to review draft documents and the comments collected electronically on them, especially in terms of resolving contentious issues.

While individual draft documents will necessarily be created by a small number of people, neglecting to have broader reviews will open the door to "false consensus," that is, agreements on important issues without critical review or insufficient consideration of possible alternatives. For metadata standards, these problems can be fatal in terms of adoption. As with most standards, if the user community does not have sufficient input into the standard, the standard is likely to be irrelevant and ignored.

A challenge in the development of written metadata standards is to generate the broadest meaningful knowledge on the scope and details of the subject and then to produce complete draft documents that can be extensively reviewed and improved. The nanomaterials WG was structured to have specific individuals tasked with producing the draft at well-defined times. This allowed for continuous progress, especially as having complete drafts exposed areas in which definitional problems or lack of clarity existed. The workshops that reviewed the drafts allowed for faster and more collective improvements.

5.6. Collaborating with existing work

This last consideration makes it imperative to involve existing work on the metadata in question. This work includes existing databases, data repositories, research projects and centers, funding organizations, industry groups, instrument manufacturers, and related standards development organizations. Working with existing groups can mean that existing metadata standards may need to be extended, changed significantly, or discarded based on additional viewpoints. Because the existing standards, even if highly informal, have resulted from investments of time and money, groups may be quite reluctant to abandon or change existing approaches. Involving such parties from the beginning is an important strategy in getting acceptance of needed changes.

For the nanomaterials metadata standards work, collaborations were set up with existing data repositories (Mills et al. 2014) (Hastings et al. 2015), standards organizations (ISO TC 229 (ISO TC 229, 2016), ASTM E56 ("ASTM E56 on Nanotechnology," n.d.), and CEN 229 ("CEN|HOME PAGE," n.d.)), national and international coordinating bodies such as the National Nanotechnology Coordination Office (U.S. NNCO), the Chinese Nanotechnology Center, and the EU Future Nano Needs project. Because workshops were held in North America, Europe, and Asia, face-to-face meetings greatly facilitated this coordination.

5.7. Project management

As with all data projects, a metadata standards project should have clear deliverables and timetables. Defining the deliverables and managing progress towards achieving them requires project management skills that are different from research management. In research, managers not only have targets, but also keep an eye out for unexpected results, that may, in fact, totally change the research goals – e.g., research fails or a new research direction is indicated. The expectation in a data project is that the data will be collected, a database will be built, and metadata standards will be developed and agreed upon. Tracking goals and progress using project management tools such as software is important.

In the majority of situations involving metadata standards, most of the work will be done by volunteers or paid personnel only partially dedicated to the project. Consequently these constraints need to be factored into the project management to keep to a realistic schedule.

A second facet of project management that is equally important is reporting requirements, whether to funding agencies or participant organizations. Metadata standards work always seems to take longer than expected, and it is important to keep participants, funders, and supporting organizations well-informed so their interest does not falter.

Another aspect of project management is to maximize the use of available funding, which, as discussed below, can be quite limited for metadata standards work. Because multiple funding sources are usually involved, most of which are not under the central control of one manager, coordination for funding travel, salaries, and meeting expenses can be difficult.

One of the leaders of the nanomaterials metadata standards project (John Rumble) has over three decades experience in organizing and leading scientific and technical metadata projects on a national and international level. This experience, combined with that of leading nanomaterials researchers and database developers, was important in keeping the project on track in terms of schedule and scope.

5.8. Longer-term support

Short-term and starting funding was discussed above, but for multi-year projects, long-term funding and long-term commitment of people is needed. An unfortunate fact is that most research funding agencies are not interested enough in the management and preservation of their research data to want to fund metadata standards work. Such work seems to have too little glamor and too low impact to interest them. The counter-example of these views – the enormous success of the crystallographic community in using metadata standards to preserve and make available virtually the entire body of crystallographic research over the last seven decades and more (Hall and McMahon, 2005) has not changed this attitude, even in this era of Big Data.

Fortunately, the enthusiasm of scientists dedicated to capturing and preserving data can prevail in the face of institutional indifference, and longer-term funding can be found in small increments to larger research funding and small individual grants.

The nanomaterials project has lasted about seven years, though funding has slowly diminished. Most of the remaining work has continued using a variety of smaller resources. Longer term support remains a critical issue.

5.9. Sustainability

A metadata standard should be designed to be used and maintained over the years, and the long-term care of such standards should be an integral part of planning and executing the project. A variety of homes are possible.

Formal standards development organizations (SDOs), such as ISO, CEN, and ASTM as mentioned above, have the infrastructure to publish, distribute (for a fee), maintain, and update standards of all types. Membership on the committees and subcommittees are controlled by those organizations (see references for details), but these groups have a long-term stability that is useful. Documents produced by a WG must be voted upon by the respective committees, which may result in changes (unintended or otherwise) to what was produced by a WG.

International scientific unions have the advantage of having an infrastructure deeply interested in a specific discipline; this, however, may be a disadvantage for metadata standards that cover multiple disciplines. The International Union of Crystallography (IUCr) (Hall and McMahon, 2005) and the International Union of Pure and Applied Chemistry (IUPAC) (IUPAC, n.d.) have maintained metadata standards on crystallographic data and chemical nomenclature respectively for decades.

Other international organizations, such as the Organization for Economic Development (OECD) (OECD WPMN, 2016) and the International Atomic Energy Agency (IAEA), have dealt with specialized metadata standards in various disciplines.

National standards development organizations are as formal as the international SDOs but usually have less rigorous procedures and more accommodating infrastructure. Their drawback is that usually participation is limited to experts from one country.

International data organizations, such as CODATA and the World Data System (WDS, n.d.), are less formal bodies yet with a long-term existence and keen interest in advancing scientific and technical data work.

Informal working groups do exist, but face problems of limited infrastructure, insufficient funding, and participation of needed expertise.

Regardless of the type of long-term sustainability path, working groups must produce sufficient documentation of their work, as well as appropriate publicity. Groups such as international data organizations and international scientific organizations are important venues for publishing reports and details of adopted metadata standards. In addition, informing the research and data communities of the standards and details of implementation is critical to adoption, and articles in archival are especially important both for immediate notification and long-term archiving of such standards.

Good metadata standards will result in changed practices in the data generation community, with more details being reported in research papers and reports. Similar changes are likely in the data community with capture of greater details and requirements for contributors to data repositories for adherence to good data and metadata practices.

The work of the CODATA WG on Nanomaterials has passed to a formal standards development organization (ASTM E56) ("ASTM E56 on Nanotechnology," n.d.), and several standards have been approved and are in the process of being approved. Certain technical areas, such as detailed description of the surface of a nano-object or the topology of a collection of nano-objects, will be addressed by the ASTM E56 and ISO TC 229. The passing of technical sustainability from a CODATA WG to a more formal body is commonplace and recognizes that such working groups start a process of deeper knowledge acquisition.

6. Challenges

Metadata standards projects face a number of practical challenges that if anticipated can be overcome with good planning and reasonable expectations. Here we identify some of these challenges and provide ideas for how to avoid or mitigate them.

Maintaining long term interest: Developing and reaching consensus on metadata standards can be difficult and takes time. Keeping the attention and interest of WG experts requires some thought and care. Strategies that have worked include creating subgroups with short-term responsibility to write a draft or identify and

define certain metadata terms with target deadlines of a few weeks or months including review and consensus development within the subgroup. Once that work is done, the subgroup can be disbanded and a new subgroup with a different mix of members can be established. Face-to-face meetings to review collectively the entire body of work developed by individual subgroups can proceed efficiently. Subgroups can be as small as two people but should be no larger than five or six.

The point is to avoid putting the burden of development on a small subset of the entire WG with the consequent result that months or even years lapse between completion of draft documents.

Long term funding: If commitments of long-term funding are not immediately available, the work can be divided into much smaller pieces based on some priority criteria and proceed step-by-step. If a serious metadata issue is being addressed by the project, then support can slowly build up over time as progress is demonstrated. It is unlikely that all involved disciplines in a multi-disciplinary project will provide support. This is especially true when one discipline is primarily a consumer of data generated by other disciplines.

Collaborating with existing work: This not only allows a project to take advantage of existing knowledge, but also helps preclude rejection of the project results for reasons such as "it is too costly or too time-consuming to redo our database structure."

Accessibility and IPR: The more formal a standard produced, i.e., acceptance and publication by an international or national SDO, the more likely that body will assert IPR privileges, demand copyright, and charge fees for the resulting document. Such arrangements need to be discussed in advance while descriptive articles in archival journals or explanatory companion documents can greatly increase the accessibility. This is especially needed if the metadata standards require the data generation community to change its reporting practices about experimental and theoretical parameters relevant to the measurement result.

Publicity: Standards are created to be used, but often considerable publicity is needed to make the community aware of the metadata standards developed. Websites, papers, talks, and articles are effective mechanisms, but care must be taken to reach the breadth of disciplines affected. Periodic articles in discipline-specific news publications are especially effective at keeping that discipline informed of progress.

Competition: In rapidly developing disciplines, the need for metadata standards may already be addressed informally or incompletely by *ad hoc* or more formal groups. Usually these groups, however, do not represent the full set of disciplines that should be involved, but at the same time, these existing efforts may be reluctant to give up their activity. Every effort should be made to persuade such group to join this new larger effort.

Face-to face meetings: As mentioned several times above, face-to-face meetings are critically important to ensure active participation, meaningful review, and real consensus. When pursuing funding, full attention should be given to supporting these meetings. To mitigate overseas travel costs, regional meetings are effective if the project leaders are able to attend all such meetings.

Unresolvable metadata problems: Our knowledge of the cause and effect of all measurable quantities is obviously incomplete, and the importance of individual independent variables is often not known. As a result, serious problems or disagreements may arise about which metadata items to include or how best to define them precisely. This is more a reflection of our lack of scientific knowledge than a flaw in the metadata standards development process. It is acceptable to explicitly indicate ambiguity or competing viewpoints, or even to point out a problem and leave it unresolved. In such cases, it is important to point out the reasons for this, which helps them to be resolved in the future.

Differences in international participation and expectations: In addition to individual scientists often being uncomfortable with the standards development process, different countries have very different cultures with respect to their participation in such work and their expectations for the results. In some countries, the scientific hierarchy is fairly rigid and participation of experts in international work may be decided by seniority or institutional affiliation rather than by expertise. The availability of travel and time support to participate in the activity may also be greatly affected by such factors. This also applies to more formal standards work where countries can have very complicated and well-defined procedures for selecting participants in SDOs.

CODATA has a long history of being a neutral body that facilitates overcoming these barriers, and CODATA Working Groups often can relax difficult rules in this area.

Acceptance of results: While standards are well understood by engineers and industry, scientists, researchers, and academia are less comfortable with them. Standards often involve taking a step backward to the

more widely-accepted discipline norms, which can cause skepticism among those accustomed to pushing the state-of-the-art. The business case of saving money and time by preserving data can be less persuasive with scientists who, due to current performance measurement systems, are often focused on the competitive pursuit of individual rather than community goals. Here the International Unions can help, especially those already involved in developing standard practices and procedures.

7. Applicability to Other Disciplines

The approach for multi-disciplinary nanomaterials standards was adapted from previous work by other CODATA groups and experience with materials and chemical data standards. For example, CODATA groups have long provided catalytic work on international collaboration on data evaluation and data standards for individual disciplines in areas such as the fundamental physical constants, thermochemical data, and rates of chemical reactions. Success in these projects led to broader collaboration, often with CODATA providing initial leadership with later transfer to ongoing oversight bodies. These broader collaborations gradually extended into multi-disciplinary efforts, such as biodiversity, astronomy (across different observation methods), and climate change. The parameters for success in those projects provided guidance for work on nanomaterials metadata standards.

There is nothing unique in any of these projects that makes the approach described above to be discipline specific. Different disciplines vary in practices such as publishing (contrast the growth of open access or prepublication archives in disciplines as different as physics, earth science, physics, and biodiversity), the maturity and acceptance of large-scale data repositories and deposition requirements (compare genome data with solid state physics), and the amount of metadata usually reported. The overall practice of science is essentially the same regardless of disciplines, being a mixture of large and small groups, international in scale, diverse in their funding sources, and increasingly multi-disciplinary.

The approach used for nanomaterials metadata standards, as described and based on other similar successful data collaborations, is quite independent of discipline and can be adapted to other multi-disciplinary metadata standards activities.

8. Conclusions

We have describe in this paper the elements of a successful multi-disciplinary metadata standards project as well as the challenges and barriers that such projects may experience. Using the recent experience in metadata standards for nanomaterials data performed by a CODATA Working Group, we have tried to point out strategies to make such projects successful and useful. To summarize that experience, we identify the critical elements that we believe make such work useful and worthwhile. These include:

- *Strong planning*
- *Active leadership*
- *Clear goals*
- *Realistic expectations for volunteer participations*
- *Inclusion of affected disciplines*
- *International involvement*

The growing importance of Big Data and Data Science in the progress of 21st century science forcefully increases the need for large, high-quality data collections. In turn, such collections are useful and accessible in the multi-disciplinary environment of today's science only if strong, robust, and internationally accepted metadata standards are available.

As the major international scientific data organization, CODATA is prepared to help meet this need.

Acknowledgements

The CODATA WG on Nanomaterials was also co-chaired by Steve Freiman and Clayton Teague, whose contributions to its success are gratefully acknowledged.

Competing Interests

The authors wish to acknowledge funding support from the International Science Council (ISC) and CODATA. The authors also acknowledge support from the FutureNanoNeeds project funded by the European Union Seventh Framework Programme (FP7-NMP-2013-LARGE-7) under grant agreement no. 604602.

References

- ANSI SDOs.** n.d. Standards Developing Organizations (SDOs) [WWW Document]. URL: https://www.standardsportal.org/usa_en/resources/sdo.aspx (accessed 9.23.18).
- ASTM E56 on Nanotechnology** [WWW Document]. n.d. URL: <https://www.astm.org/COMMITTEE/E56.htm> (accessed 8.31.17).
- ASTM International.** n.d. ASTM International – Standards Worldwide [WWW Document]. URL: <https://www.astm.org/> (accessed 9.23.18).
- Boverhof, DR, Bramante, CM, Butala, JH, Clancy, SF, Lafranconi, M, West, J and Gordon, SC.** 2015. Comparative assessment of nanomaterial definitions and safety evaluation considerations. *Regul. Toxicol. Pharmacol*, 73: 137–150. DOI: <https://doi.org/10.1016/j.yrtph.2015.06.001>
- Bruno, I and Frey, JG.** 2017. Connecting chemistry with global challenges through data standards. *Chem. Int*, 39: 5–8. DOI: <https://doi.org/10.1016/j.yrtph.2015.06.001>
- CEN.** n.d. Home [WWW Document]. URL: <https://www.cen.eu/Pages/default.aspx> (accessed 9.23.18).
- CEN|HOME PAGE** [WWW Document]. n.d. URL: <https://www.cen.org/index.cfm?> (accessed 9.23.18).
- CODATA.** 2018. CODATA/VAMAS Joint Working Group on the Description of Nanomaterials – CODATA [WWW Document]. URL: <http://www.codata.org/nanomaterials> (accessed 8.7.18).
- CODATA.** n.d. CODATA, The Committee on Data for Science and Technology [WWW Document]. URL: <http://www.codata.org/> (accessed 9.23.18).
- Dale, AL, Casman, EA, Lowry, GV, Lead, JR, Viparelli, E and Baalousha, M.** 2015a. Modeling nanomaterial environmental fate in aquatic systems. ACS Publications.
- Dale, AL, Lowry, GV and Casman, EA.** 2015b. Stream Dynamics and Chemical Transformations Control the Environmental Fate of Silver and Zinc Oxide Nanoparticles in a Watershed-Scale Model. *Environ. Sci. Technol*, 49: 7285–7293. DOI: <https://doi.org/10.1021/acs.est.5b01205>
- Hall, SR and McMahan, B.** 2005. International Tables for Crystallography, Definition and Exchange of Crystallographic Data. Springer Science & Business Media.
- Hanisch, RJ.** 2015. The development, deployment, and impact of the virtual observatory, Part II. *Astron. Comput*, 11: 73–73. DOI: <https://doi.org/10.1016/j.ascom.2015.04.001>
- Hastings, J, Jeliaskova, N, Owen, G, Tsiliki, G, Munteanu, CR, Steinbeck, C and Willighagen, E.** 2015. eNanoMapper: Harnessing ontologies to enable data integration for nanomaterial risk assessment. *J. Biomed. Semant*, 6: 1. DOI: <https://doi.org/10.1186/s13326-015-0005-5>
- Hill, DP, Adams, N, Bada, M, Batchelor, C, Berardini, TZ, Dietze, H, Drabkin, HJ, Ennis, M, Foulger, RE and Harris, MA.** 2013. Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology. *BMC Genomics*, 14: 513. DOI: <https://doi.org/10.1186/1471-2164-14-513>
- ICSU.** 2018. Home [WWW Document]. *Int. Sci. Counc.* URL: <https://council.science/> (accessed 9.23.18).
- IEC.** n.d. Welcome to the IEC – International Electrotechnical Commission [WWW Document]. URL: <https://www.iec.ch/> (accessed 9.23.18).
- ISO – International Organization for Standardization** [WWW Document]. n.d. URL: <https://www.iso.org/home.html> (accessed 9.23.18).
- ISO TC 229.** 2016. ISO Technical Committee 229 Nanotechnologies [WWW Document]. *ISO Tech. Comm. 229 Nanotechnologies*. URL: http://www.iso.org/iso/iso_technical_committee?commid=381983.
- IUPAC.** n.d. International Union of Pure and Applied Chemistry [WWW Document]. *IUPAC Int. Union Pure Appl. Chem.* URL: <https://iupac.org/> (accessed 9.23.18).
- Karcher, S, Willighagen, EL, Rumble, J, Ehrhart, F, Evelo, CT, Fritts, M, Gaheen, S, Harper, SL, Hoover, MD and Jeliaskova, N.** 2018. Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations. *NanoImpact*, 9: 85–101. DOI: <https://doi.org/10.1016/j.impact.2017.11.002>
- Marchese-Robinson, RL, Lynch, I, Peijnenburg, W, Rumble, J, Klaessig, F, Marquardt, C, Rauscher, H, Puzyn, T, Purian, R, Åberg, C, Harper, S, et al.** 2016. How should the completeness and quality of curated nanomaterial data be evaluated? *Nanoscale*, 8: 9919–9943. DOI: <https://doi.org/10.1039/C5NR08944A>
- Mills, KC, Murry, D, Guzan, KA and Ostraat, ML.** 2014. Nanomaterial registry: Database that captures the minimal information about nanomaterial physico-chemical characteristics. *J. Nanoparticle Res*, 16: 2219. DOI: <https://doi.org/10.1007/s11051-013-2219-8>
- OECD WPMN.** 2016. OECD Working Party on Manufactured Nanomaterials [WWW Document]. *Saf. Nanomater.* URL: <http://www.oecd.org/science/nanosafety/>.

- Quinn, PJ, Barnes, DG, Csabai, I, Cui, C, Genova, F, Hanisch, B, Kembhavi, A, Kim, SC, Lawrence, A and Malkov, O.** 2004. The International Virtual Observatory Alliance: Recent technical developments and the road ahead. In: *Optimizing Scientific Return for Astronomy through Information Technologies. International Society for Optics and Photonics*, 137–146. DOI: <https://doi.org/10.1117/12.551247>
- Roco, MC.** 2011. The Long View of Nanotechnology Development: The National Nanotechnology Initiative at 10 Years. In: Roco, MC, Hersam, MC and Mirkin, CA (eds.), *Nanotechnology Research Directions for Societal Needs in 2020: Retrospective and Outlook*, 1–28. Dordrecht, Netherlands: Springer. DOI: https://doi.org/10.1007/978-94-007-1168-6_1
- Rumble, JR.** 2017. Accessing Materials Data: Challenges and Directions in the Digital Era. *Integrating Mater. Manuf. Innov*, 6: 172–186. DOI: <https://doi.org/10.1007/s40192-017-0095-2>
- Sayle, R.** 2009. Foreign language translation of chemical nomenclature by computer. *J. Chem. Inf. Model*, 49: 519–530. DOI: <https://doi.org/10.1021/ci800243w>
- WDS.** n.d. Introduction – World Data System: Trusted Data Services for Global Science [WWW Document]. URL: <https://www.icsu-wds.org/organization> (accessed 9.23.18).

How to cite this article: Rumble, J, Broome, J and Hodson, S. 2019. Building an International Consensus on Multi-Disciplinary Metadata Standards: A CODATA Case History in Nanotechnology. *Data Science Journal*, 18: 12, pp. 1–11. DOI: <https://doi.org/10.5334/dsj-2019-012>

Submitted: 23 October 2018 **Accepted:** 14 March 2019 **Published:** 09 April 2019

Copyright: © 2019 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 