Practical Data Sharing: The Nanomaterial Registry

M. L. Ostraat*, D. Murry**, K. A. Guzan*** and K. C. Mills****

*RTI International, 3040 Cornwallis Road, RTP, NC, 27709, USA, mostraat@rti.org
**RTI International, 3040 Cornwallis Road, RTP, NC, 27709, USA, dmurry@rti.org
***RTI International, 3040 Cornwallis Road, RTP, NC, 27709, USA, kguzan@rti.org
****RTI International, 3040 Cornwallis Road, RTP, NC, 27709, USA, kmills@rti.org

ABSTRACT

The Nanomaterial Registry is a publicly available, authoritative, central location for nanomaterial data sharing that was created in order to accelerate nanomaterial product innovations as well as discoveries on nanomaterial biological and environmental implications (www.nanomaterialregistry.org). The Registry is the first community resource for collaborative data analysis based on a minimal information about nanomaterials (MIAN).

A rigorous curation process is used to systematically organize and archive all nanomaterial records. A record is defined by a unique nanomaterial sample, including metadata, such as the circumstances under which the sample was characterized, hyperlinks to the original publication(s) of the data, manufacturer information, and characterization protocols. Compliance levels are calculated for all characterization data to indicate the quantity and quality of information available in each record.

Keywords: registry, curation, minimal information standards, nanomaterial

1 INTRODUCTION

The central issues hindering community data collaboration have been identified by the National Nanotechnology Initiative's (NNI's) Nanotechnology Signature Initiative (NSI) for a Nanotechnology Knowledge Infrastructure (NKI): (1) the lack of standardized data management procedures and (2) the lack of public access to a sustainable data sharing system. As a result, the National Institutes of Health (NIH) have, through the Nanomaterial Registry, sponsored the development of a mimimal information about nanomaterials (MIAN) for characterization as well as the building of the database to house this information. The Nanomaterial Registry is being developed by RTI and is the first publicly available registry for nanomaterial data that (1) uses an applied MIAN, (2) provides scores for data quality based on the MIAN, and (3) provides the broad nanomaterial community with a central and authoritative repository for the purposes of regulatory decision making, knowledge gap evaluations, and downstream modeling analysis. The Nanomaterial Registry, in its third year, offers practical solutions to both of the NKI's central data collaboration issues through evergreen data curation procedures, uniform structuring of data, and a public website which allows for collaborative analysis by the community (www.nanomaterialregistry.org).

The data in the Nanomaterial Registry are fully curated using the MIAN and a controlled vocabulary so that characterizations and studies from different arenas of nanomaterial research are semantically interoperable (figure 1). The Nanomaterial Registry curates from several different sources of nanomaterial characterization data and biological and environmental study data. The majority of these data are publicly available and drawn from publicly available databases. The remaining data are gathered from

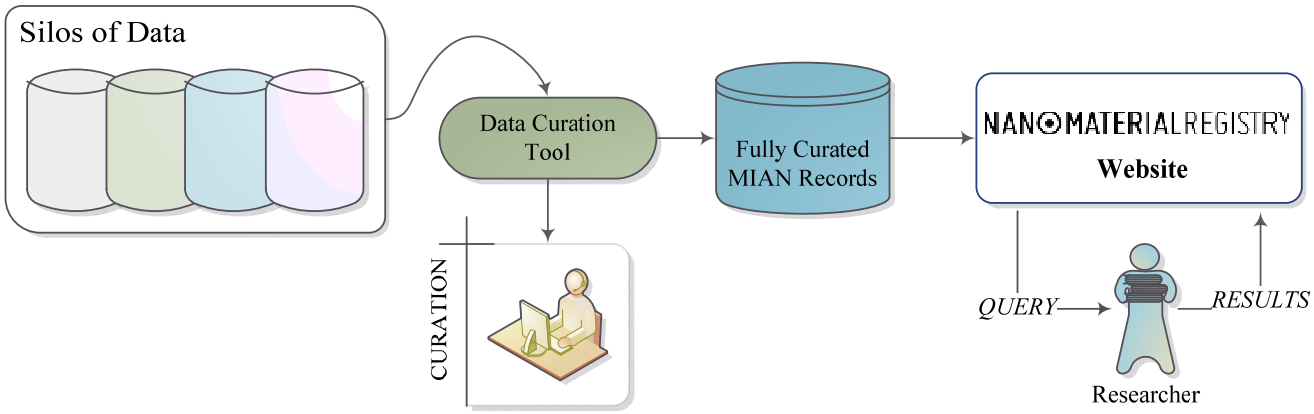

Figure 1. The Nanomaterial Registry's conceptual data flow from data sources, to curation, to publishing in the Nanomaterial Registry database, and availability on the Registry site.

strategic partner and standards organizations, possessing rich amounts of nanomaterial data. In order to communicate the quality and quantity of the data curated, compliance algorithms are applied to the data, which are then reflected on the Nanomaterial Registry's website via compliance levels of merit, bronze, silver, or gold.

The MIAN covers (a) 12 physico-chemical characteristics (PCC), the techniques, instrument parameters, and scientific best practice questions for each measurement; (b) metadata around the source of the data and the time point at which the data were collected; and (c) description categories on any biological or environmental studies in which the nanomaterials were used. While the scope of the Registry is that of housing minimal information, this is the first applied MIAN in the nanomaterial community. So, it is understood that as discoveries are made on nanomaterials and their implications, the list of essential characteristics may need to be modified.

The Nanomaterial Registry was built and developed with counsel from an advisory board consisting of representatives from various stakeholder groups of nanotechnology: regulatory, academia, industry, and government. Through their advisement the Nanomaterial Registry was positioned to be a central repository of authoritative data that not only spans the broad nanomaterial communty, but is also deep in information on the nanomaterials of highest concern to the community [1].

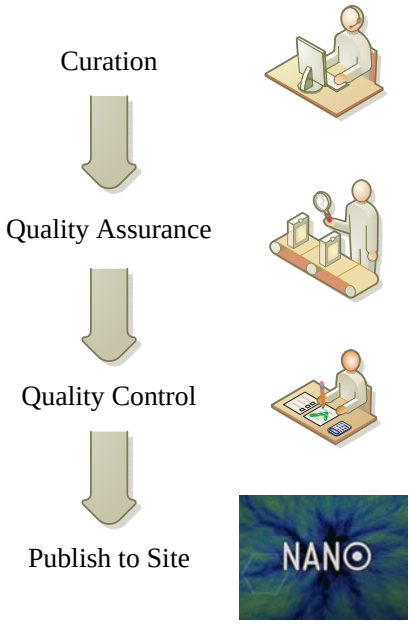

Figure 2. Nanomaterial records are promoted through the stages of curation before being published to the Nanomaterial Registry website.

2 DATA CURATION PROCESS

Data curation for the Nanomaterial Registry is performed using a web-based tool designed for the purposes of process efficiency, error mitigation, and data validation, which have all been recognized as critical elements to sustainable data curation and archival practices [2-5]. The tool uses a promotion process to curate and verify the correctness of the data before they are published to the Nanomaterial Registry website (figure 2). Once curated from the data source, a nanomaterial record passes to the quality assurance (QA) stage, where it is reviewed for grammar and punctuation usage and completeness of information. After QA, the nanomaterial record is promoted to quality control (QC) and is reviewed for appropriate scientific interpretation and, once approved, is published.

Many fields for curation in the Nanomaterial Registry's curation tool have been populated by pick lists designed from the MIAN and controlled vocabulary. These picklists eliminated the errors associated with free text fields and reduced the time used in all phases of the curation process. The tool was designed to also provide curators with a decision tree-like procedure when curating a record, which simplified the curation phase.

3 DATA RECORD STRUCTURE

Each record in the Nanomaterial Registry is given a unique identifier. These identifiers are numeric and assigned sequentially by the order in which the records are curated. The use of human-readable nanomaterial identifers is a community-wide goal and there have been methods proposed, such as Thomas et al [6] and Gentleman and Chan [7]. A system will be adopted by the Nanomaterial Registry once it has become standard.

The highest level of data in a nanomaterial record contains information on the institutional source of the data and information on any publication in which the data were published, see figure 3. This information not only validates the data by providing reference to the original sources, but also connects multiple data records acquired from the same data source and/or the same publication and study.

The structure for research data in the Nanomaterial Registry was designed in a hierarchical format using the concept of instance of characterization (IOC) as the root sorting method for data sets [8,9]. Within each nanomaterial record, could reside several IOC data sets, which are separated based on time points and physical state of the nanomaterial sample at the time of characterization and/or biological or environmental study. Examples of information contained at the IOC level of data are the manufacturer and product ID, the laboratory in which the study was performed and any information on the synthesis method of th nanomaterial. Beneath each IOC, are curated the MIAN PCC infornation, and the MIAN for the biological and environmental studies.

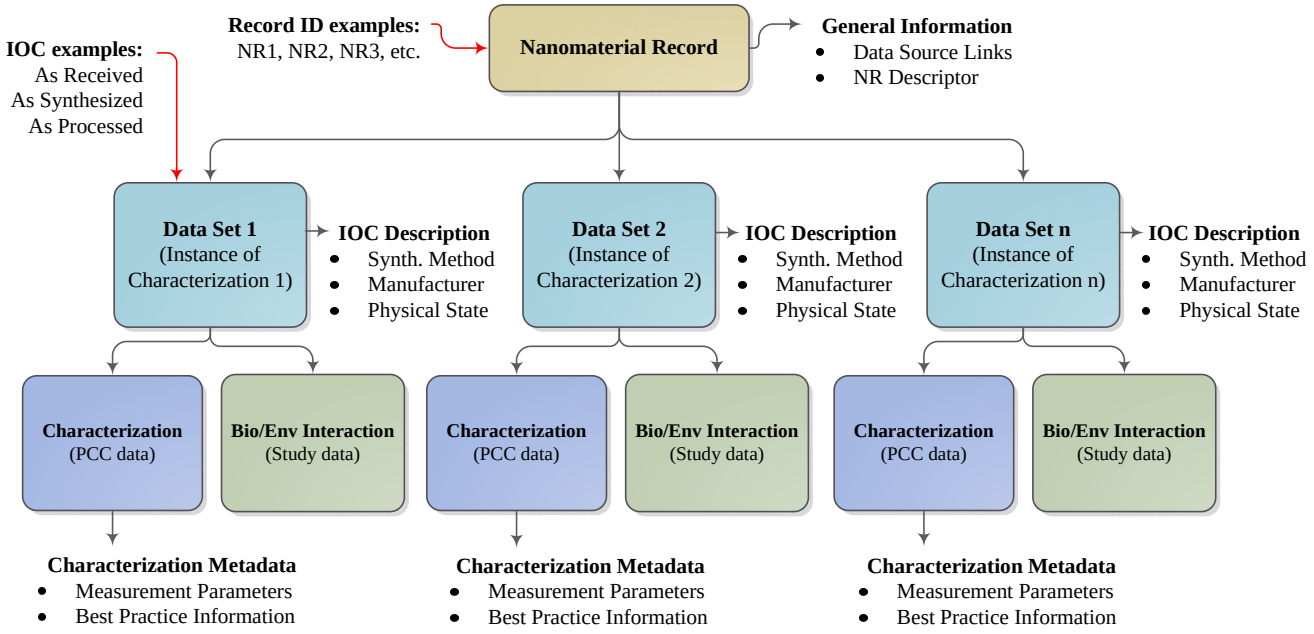

Figure 3. Each nanomaterial record is given a unique identifier. The structure of a nanomaterial record in the Nanomaterial Registry database is rooted on information pertaining to the nanomaterial's instance of characterization (IOC). Within each IOC are the physico-chemical characterization data and/or information on biological and environmental studies in which the nanomaterial was tested.

Each PCC measurement is accompanied by metadata of the measurment uncertainty, the technique and analytical instrument names, the protocol and parameters crucial to the technique, and answers to best practice questions, such as, "How many replicates were performed?" and "Was the instrument within calibration?"

In order to provide users with the most opportunity to utilize the curated data in downstream analyses, all information in the Nanomateial Registry database is parsed as much as possible, using standard vocabulary in many cases [10,11]. The use of standard vocabulary and a controlled vocabulary throughout the Nanomaterial Registry becomes essential in a data repository housing information from disparate resources. The implementation of semantic interoperability within the data is another way the Nanomaterial Registry provides rich data for downstream analyses.

4 CONCLUSIONS

As the population of data in the Nanomaterial Registry grows, so too does the need for increased efficiency in data curation. In preparation for this, the Nanomaterial Registry is currently pursuing, in cooperation with other community groups, a standardized data format for nanomaterial data, called ISA-Tab-Nano [12]. This format could be used in the laboratory for daily record keeping and would also be interoperable with the Nanomaterial Registry. Data could be uploaded by the researcher directly into the curation process, thus reducing the burden of curators to parse information. A data upload tool would be available at the Nanomaterial Registry website to facilitate this. A standard data format is also adventageous for data export from the Nanomaterial Registry's website. Exported data could be used for advanced analyses and modeling.

The data records curated and housed by the Nanomaterial Registry can be found by using the search or browse features. Nanomaterial records of interest can be compared side by side directly on the website. Also, a simple rule-based matching has been applied to the nanomaterials which aids the user in accessing more nanomaterials similar to any individual record of interest.

The Nanomaterial Registry expects to evolve as the community makes discoveries about nanomaterials. With the ever-growing amount of data available, users of the Nanomaterial Registry will be able to find new trends and relationships in nanomaterial PCCs and system implications. Through their published discoveries, the Nanomaterial Registry will be updated and additional findings and knowledge on the implications of nanomaterials will be possible.

5 ACKNOWLEDGEMENTS

RTI acknowledges funding for this project, under contract HHSN268201000022C, from the following institutions within the National Institutes of Health (NIH): the National Institute of Biomedical Imaging and

Bioengineering (NIBIB), the National Institute of Environmental Health and Sciences (NIEHS), and the National Cancer Institute (NCI).

REFERENCES

[1] OECD's Working Party on Manufactured Nanomaterials. (November 2012). List of Representative Manufactured Nanomaterials for Testing. Retrieved from http://www.oecd.org/env/ehs/nanosafety/List_of_representative_MN_for_testing_Nov_2012.pdf

[2] Borgman, C. L. (2012), The conundrum of sharing research data. J. Am. Soc. Inf. Sci., 63(6):1059-1078, 2012. doi: 10.1002/asi.22634

[3] Hirschman, J.; Berardini, T. Z.; Drabkin, H. J. and Howe, D. (2010), A MOD(ern) perspective on literature curation. Mol. Genet. Genomics 283, 415-425.

[4] Green, A.; Macdonald, S. and Rice, R. (2009), Policy-making for Research Data in Repositories: A Guide. Data Information Specialists Committee – UK, version 1.2. Available from: http://www.disc-uk.org/docs/guide.pdf

[5] Ogburn, J. L. (2010), "The Imperative for Data Curation," The Johns Hopkins University Press (Retrieved from Project MUSE database), portal: Libraries and the Academy 10:2, 241-246.

[6] Thomas, D. G.; Chikkagoudar, S.; Chappell, A. R. and Baker, N. A. (2012), Annotating the structure and components of a nanoparticle formulation using computable string expressions," Bioinformatics and Biomedicine Workshops, 2012 IEEE International Conference on, 889-894, 4-7 Oct. 2012.

[7] Gentleman, D. J. and Chan, W. C. W. (2009), A systematic nomenclature for codifying engineered nanostructures. Small, 5, 426-431. doi: 10.1002/smll.200800490

[8] Sayes, C. M. and Warheit, D. B. (2009), Characterization of nanomaterials for toxicity assessment," Wiley Interdiciplinary Reviews: Nanomedicine and Nanobiotechnology 1:6, 660-670.

[9] Sayes, C. and Berg, J. M. (2011), "Physical and Chemical Properties of Nanomaterials Relevant to Toxicological Studies, In Vitro and In Vivo." Chap. 30 in Nanobiomaterials Handbook, ed: B. Sitharaman, CRC Press.

[10] ISO/TR 13014:2012: Nanotechnologies – Guidance on physico-chemical characterization of engineered nanoscale mateials for toxicologic assessment. International Organization for Standardization, Geneva, Switzerland.

[11] ISO/TR 11360:2010: Nanotechnologies – Methodology for the classification and categorization of nanomaterials. International Organization for Standardization, Geneva, Switzerland.

[12] ASTM WK28974, E2909-2013, "New Specification for a Standard File Format for the Submission and Exchange of Data on Nanomaterials and Characterizations," ASTM International, West Conshohocken, PA, 2013. Available from: http://www.astm.org/DATABASE.CART/WORKITEMS/WK28974.htm