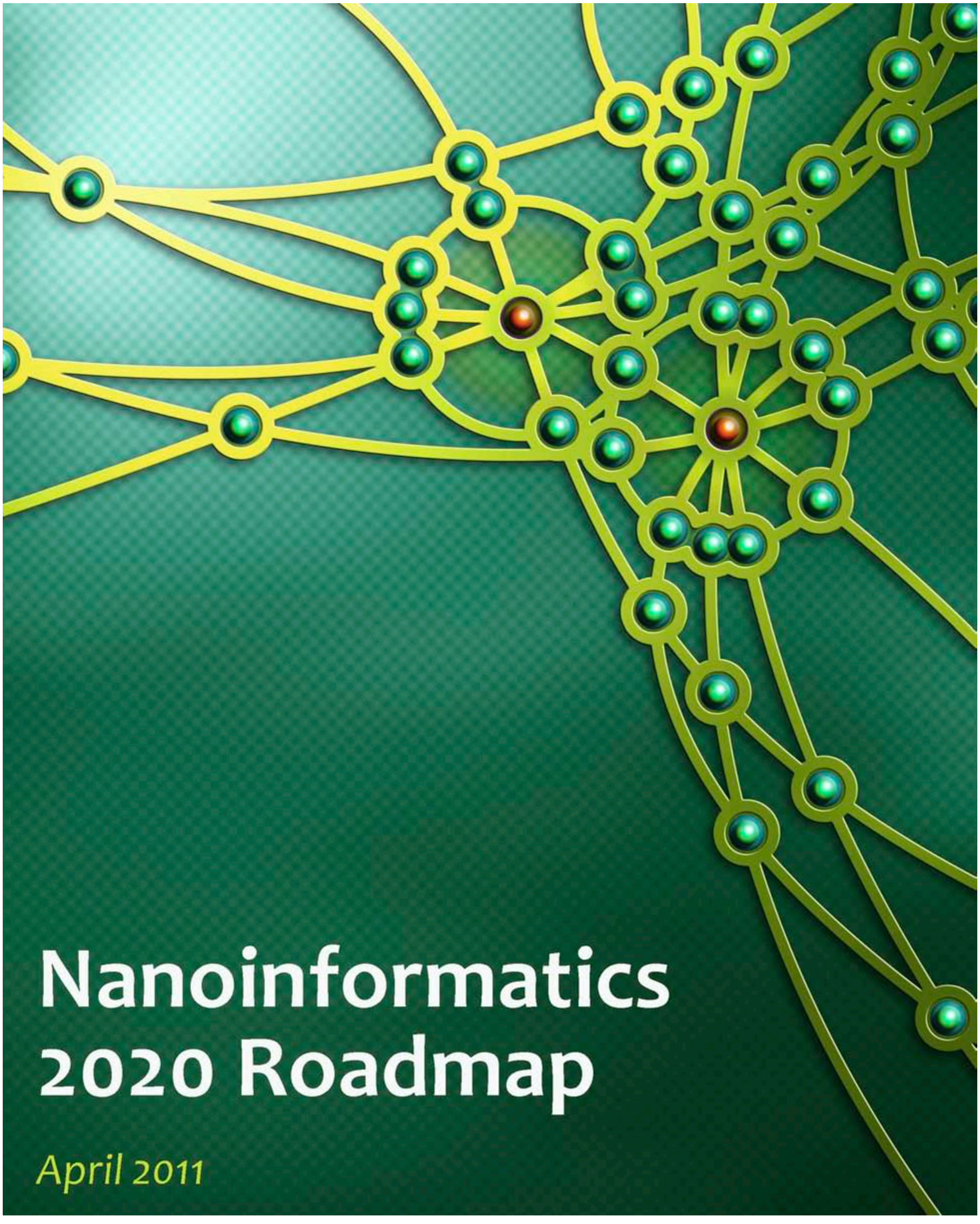


Nanoinformatics 2020 Roadmap

April 2011

Nanoinformatics 2020 Roadmap

April 2011

For more information and to download the PDF, please visit

http://nanotechinformatics.org/nanoinformatics/index.php/Main_Page

Published by the National Nanomanufacturing Network

Amherst, MA 01003

http://www.internano.org/

DOI: 10.4053/rp001-110413

This Roadmap was prepared in concert with Nanoinformatics 2010 workshop organizers and participants, with substantial input from:

Diana de la Iglesia, Universidad Politécnica de Madrid
Stacey Harper, Oregon State University
Mark D. Hoover, National Institute for Occupational Safety and Health
Fred Klaessig, Pennsylvania Bio Nano Systems
Phil Lippel, Consultant
Bettye Maddux, SNNI/ONAMI
Jeff Morse, National Nanomanufacturing Network
André Nel, University of California Los Angeles
Krishna Rajan, CoSMIC -- Iowa State University
Rebecca Reznik-Zellen, National Nanomanufacturing Network
Mark Tuominen, National Nanomanufacturing Network

Nanoinformatics 2010 was made possible by the National Science Foundation through grant number CMMI 0531171.

Nanoinformatics 2010 Organizing Committee

Nathan Baker, Pacific Northwest National Laboratory
Anne Chaka, National Institute of Standards & Technology
Yoram Cohen, University of California, Los Angeles
Vicki Colvin, Rice University
Martin Fritts, Nanotechnology Characterization Laboratory
Charles L. Geraci, National Institute for Occupational Safety and Health
Mark D. Hoover, National Institute for Occupational Safety and Health
Sharon Ku, National Institutes of Health
Kristen Kulinowski, Rice University
Phil Lippel, Consultant
James Luo, National Institutes of Health
Michael McLennan, Purdue University
Jeff Morse, National Nanomanufacturing Network
Michele Ostraat, RTI International
Krishna Rajan, CoSMIC -- Iowa State University
Rebecca Reznik-Zellen, National Nanomanufacturing Network
Peter Schad, RTI International
Mark Tuominen, National Nanomanufacturing Network

NANOINFORMATICS 2020 ROADMAP

EXECUTIVE SUMMARY

Nanoinformatics encompasses the acquisition of information relevant to nanotechnology and the development of tools for using that information efficiently. The Nanoinformatics 2020 Roadmap is the first broad-based community effort to articulate the comprehensive needs and goals in nanoinformatics. It is based in part on *Nanoinformatics 2010: A Collaborative Roadmapping Workshop*, a meeting organized by experts from the community of practice and held November 3-5, 2010. This effort responds to the call for roadmaps in the National Nanotechnology Initiative Strategic Plan.

The Nanoinformatics Roadmap is steered by a coalition of nanotechnology and informatics practitioners who foresee significant benefits emerging from strategic multidisciplinary efforts in nanoinformatics. The Roadmap serves to inform the broader nanotechnology community of the significant value informatics can add to ongoing research and development efforts. It identifies key priorities and new opportunities. It is also intended to stimulate contributions from experts in either nanotechnology or informatics regarding possibilities not foreseen by the initial members of the community of practice.

Nanoinformatics has the potential to introduce transformative new approaches to scientific discovery, fundamental research, product innovation, sustainable manufacturing, and safety to people and the environment. Nanotechnology is widely viewed as a broad-impact technology which can contribute to improved products and manufacturing processes across diverse sectors of commerce including healthcare; water and energy; transportation; defense and security; environmental remediation and environmentally friendly manufacturing; and food safety, production, and packaging. With such broad applicability comes an obligation to conform with regulatory and trade regimens, and to satisfy the expectations of heightened public scrutiny. The community's ability to meet these obligations and realize the enormous potential impact of nanotechnology depends on the acquisition, processing, and sharing of vast amounts of data. It should not be limited by a lack of suitable nanoinformatics tools; scattered, poorly organized, and uncoordinated data repositories; or barriers to data interoperability and resource pooling. An overarching nanoinformatics strategy can avoid these pitfalls by supporting coordinated nanoinformatics practices that accelerate progress in nanotechnology research, development, and manufacturing. It can support environmental, health, and safety activities (EHS) and make ongoing projects throughout the nanotechnology R&D enterprise more efficient and cost effective.

The Nanoinformatics 2020 Roadmap identifies the current stakeholders, projects, needs, capabilities, and connections that will define a successful nanoinformatics enterprise, and outlines plans for developing them. The implementation plans in the Roadmap incorporate a decade-long vision and pathway, providing a realistic timeframe to establish an effective system of nanoinformatics data, tools, and infrastructure. Such a program will enable the community to improve and "travel" on the road to understanding, development, and beneficial application of nanotechnology.

Representatives of nanoinformatics efforts occurring prior to *Nanoinformatics 2010* were vital participants at the workshop. This Roadmap builds on their foundational activities. These include several nanoinformatics projects that emerged within distinct domains of nanotechnology R&D, as well as collaborative, cross-institution, and cross-sector initiatives that developed within the last five years. These pioneering experiences provided a concrete basis for exploring the drivers for and barriers to a stronger, more effective system of nanoinformatics. The discussion, identification, and articulation of key concepts for such a system were arranged around three main themes:

- data collection and curation;
- tools for innovation, analysis and simulations; and
- data accessibility and information sharing.

The workshop participants were able to develop a broader understanding of the gaps that exist in nanoinformatics within the three main workshop themes. They also identified several cross-cutting issues which transcend the original thematic architecture and must be addressed before significant, community-wide adoption of best practices can be achieved:

- Nanoinformatics is emerging as a vital part of the research and development toolkit;
- Coordination and incentive are essential for successful nanoinformatics;
- Standards for data documentation are critical; and
- Successful informatics is a techno-social issue.

The need for focused efforts to address these issues is a central finding of the workshop, and is discussed in more detail in the section "Cross-Cutting Issues for Nanoinformatics."

The principal outcome of the workshop is this Roadmap. It is intended to become a living document, collectively steered by the nanotechnology community and open to additional input from existing or new community members. It strives to:

- Articulate the expected outcomes and impacts of a coordinated nanoinformatics effort;
- Address the drivers for and problems facing nanoinformatics research (such as the diversity of nanomaterials property data, gaps within that data, and differing needs and practices regarding data provenance, source citation, and confidentiality)
- Identify the technological dimensions of nanoinformatics, as well as the scholarly communication practices and information standards that are required for a nanoinformatics infrastructure to be efficient and to have a broad impact; and
- Provide a blueprint and serve as a benchmark for community-wide action.

Additionally, the Roadmap identifies short-term actions proposed by the 2010 workshop participants as building blocks for the robust system of nanoinformatics they envision. Specifically, it proposes seven pilot projects addressing objectives articulated at the workshop in the areas of community engagement, metadata and standards, and tool development:

- Nanomaterials data consortium;
- Nanomaterials data gaps workshops;
- Meta-ontology;
- Minimum information recommendations;
- Metacrawler;
- Nano structure-activity relationship education and dissemination; and
- Simulation challenge.

These fast-moving pilots—to be conducted in one to five years—are designed to offer catalytic solutions to technical and cultural barriers of practice; to execute proof-of-concept projects that will inform more systematic development of tools and standards; and to nurture the spirit of collaboration and cooperation that is evident among the nanoinformatics community of experts. Pilot projects will leverage already-funded cooperative activities to the maximum extent possible. Where no suitable activities are currently underway, new resources will be needed to support pilot activity.

The pilot projects represent a community-driven contribution towards the Nanoinformatics Roadmap goals of stimulating coordinated informatics activities and advocating for the integration of nanoinformatics into routine workflows across research environments. The development and widespread adoption of advanced nanoinformatics capabilities that accelerate responsible research, development, and deployment of nanotechnology–driven by the expertise and momentum of the R&D community and enabled by the agencies that support it—is the ultimate goal of the Nanoinformatics 2020 Roadmap.

1. WHAT IS NANOINFORMATICS?

1.1 WORKING DEFINITION

Nanoinformatics is the science and practice of determining which information is relevant to the nanoscale science and engineering community, and then developing and implementing effective mechanisms for collecting, validating, storing, sharing, analyzing, modeling, and applying that information.

- Nanoinformatics is necessary for intelligent development and comparative characterization of nanomaterials, for design and use of optimized nanodevices and nanosystems, for development of advanced instrumentation and manufacturing processes, and for assurance of occupational and environmental safety and health.

- Nanoinformatics also involves the utilization of networked communication tools to launch and support efficient communities of practice.

- Nanoinformatics also fosters efficient scientific discovery through data mining and machine learning.

1.2 NANOINFORMATICS BACKGROUND

Speaking broadly, *informatics* is the application of information and computer science methods for collecting, analyzing, and applying information. "X-informatics" has become the default descriptor for the application of such methods to a set of problems within a specific field or discipline,[1] such as bioinformatics in biology or ecoinformatics in ecology. Similarly, *computational science* is the use of high-powered computing and sophisticated algorithms to posit and solve problems. "Computational X" refers to the use of computational methods within a specific field or discipline (e.g., computational astronomy or computational geology). *Nanotechnology* is the umbrella term for science, engineering, and technology at the nanoscale (approximately 1 to 100 nanometers). Nanotechnology encompasses synthesizing, imaging, measuring, modeling, and manipulating matter at the nanoscale in order to understand and control materials properties. Thus, *nanoinform*atics is a systematic methodology to collect, organize, validate, store, share, model, and analyze data involved with nanotechnology processes and materials for the purpose of extracting useful information relevant to the nanoscale science and engineering communities. *Computational nanoscience*, conceived as falling within the broader term nanoinformatics, includes the development and application of the critical tools needed for simulations, computations, and predictive modeling of nanomaterials, nanoscale devices, and nanosystems.

In the last two decades, large-scale data explorations have begun to combine data-driven experimental and computational science with informatics methods utilizing massive computing networks, cutting-edge information science tools, and social networking technologies; these projects herald the beginning of the age of *e-Science*[2]. Popular examples of e-Science such as the Human Genome Project[3] and the Sloan Digital Sky Survey[4] demonstrate how computational sophistication and the coordination of domain expertise can be harmonized to address grand scientific challenges. Nanoinformatics-- the application of the e-science paradigm to nanoscale science and engineering-- targets challenges in the application of nanotechnology for the benefit of society.

1.3 VISION FOR NANOINFORMATICS

[1] Hey T. (2010) The Fourth Paradigm. [plenary] 31st Annual IATUL Conference, West Lafayette, IN, June 20 – 24, 2010.
[2] Hey T, et. al. eds. (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. 284 pp.
[3] http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
[4] http://www.sdss.org/

Informatics catalyzes the efficiency of scientific and industrial workflows across all nanotechnology domains and areas of application, including fundamental research, product innovation and manufacturing, and EHS practices.

Nanoinformatics for the workflow of research

Implemented at the heart of the research and investigation, nanoinformatics would allow researchers to leverage the findings of other efforts in support of their own investigations and to broaden the impact of their research. The traditional research lifecycle is oriented around pre-and post-publication milestones of information collection, preparation, analysis, and dissemination, all of which have conventionally taken place on distinct and unintegrated platforms, or silos. Developments in computational technology and networked communication allow communities of practice to integrate the silos of different domains and at different stages of investigation to achieve outcomes more efficiently and to deepen their impact throughout the domain.

For example, experimentation can be conducted using high-throughput tools and minimum data standards to capture large-scale, standards-compliant data sets. Using mapping, visualization, and advanced analytical tools, a researcher may uncover important information which points research in new directions. Such cyber-enabled discoveries can quickly advance the exploration and application of systems too complex to be understood solely from first-principles science. Similarly, modeling and simulation efforts, subject to robust validation and verification, can provide information that complements experimental data and motivates subsequent research. Data can be made publicly available prior to or in conjunction with publication through established data repositories and can benefit from the use of standard attribution and identification mechanisms such as digital object identifiers (DOIs) for data sets. Data sets and supplementary files can be mined, along with peer-reviewed literature, for trends and gaps. Semantic search algorithms, federated data, and ontologies all contribute to the discoverability and reuse of data, which seeds future investigations. Increasingly sophisticated tools for networked communication and collaboration tie all of the components of the life cycle together. (See Figure 1.)

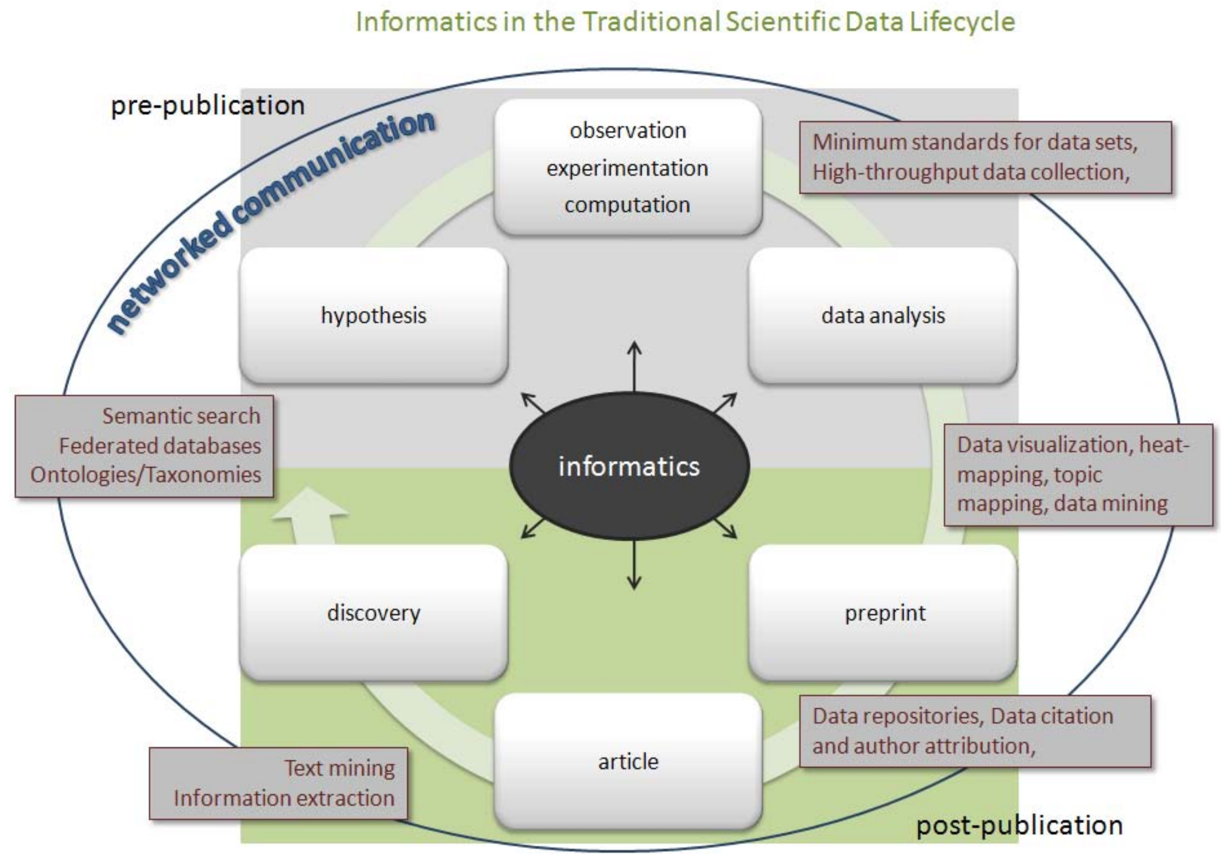


Figure1[5]

The utilization of appropriate nanoinformatics mechanisms throughout the research cycle will ultimately lead to the efficient advancement of nanotechnology research and the commercialization of nanotechnology-enabled products and systems.

Nanoinformatics for the workflow of product innovation and manufacturing

Another area poised to greatly benefit from nanoinformatics is product development and manufacturing. Although distinct from the activities involved in fundamental research, the industrial workflow associated with nanotechnology commercialization shares several common attributes. Data-driven activities in industry accelerate product design, product performance, reliability, manufacturing design, logistics, quality control, safety, and the business model. Indeed, there are many opportunities for productive synergy between the nanoinformatics activities of scholarly research and those of industrial development. In the present day, new technology development relies heavily on the data and information provided by fundamental research activities. Access to more complete data sets than the limited representative examples appearing in the published scientific literature could enable feasibility

[5] Figure adapted from Gold A. 2007. Cyberinfrastructure, Data, and Libraries Pt. 1. D-Lib Magazine 13 9/10. doi: 10.1045/september20september-gold-pt1

studies and design activities in industry. Nanoinformatics tools can streamline workflow activities and reduce the time to market. Similarly the use of nanoinformatics tools leading to the identification of data or modeling gaps in industry can drive new fundamental research activities to fill those needs.

From an industrial perspective, the value-added of each process step in a manufacturing chain of activity (the value chain) is better controlled and optimized when it is a data-rich activity. The design and manufacturing of products that contain nanomaterials require specific properties data that affect the overall performance. The raw materials, which in some cases are themselves nanomaterials, should arrive with materials certification data that describe their structure and properties. Each subsequent manufacturing process step has inputs that affect the resulting output structure and properties. Nanoinformatics-based simulation and modeling in combination with metrological process data can enable accurate process control and optimization. By performing a sensitivity analysis using known input distributions, variabilities, and process-property relationships, manufacturing reproducibility can be predicted and design margins can be established, making nanomanufacturing process optimization and adaptable manufacturing viable. In the manufacturing stream there should be, ideally, a stream of process and characterization data providing provenance data for standardization, extensibility, and new manufacturing innovations.

Nanoinformatics for the workflow of EHS practices

Additional data-rich measures performed concurrently with production data activities can support good EHS practices through the entire product lifecycle. Efforts at national laboratories, academic research centers, and forward-looking industrial sites strive to collect data that identify any potential risk of nanomaterials to health or the environment, perform site testing to evaluate exposure potential, and establish best practices for worker safety. Making the results of these efforts available broadly through information repositories will provide pathways for innovative, safe, and sustainable manufacturing of nanotechnology-enabled products.

These envisioned modifications to research, production, and EHS practices are profound. Such broad-based change cannot be implemented immediately; current nanoinformatics tools are nascent at best, and cultural shifts need time to percolate. This roadmap therefore recommends a graded introduction of nanoinformatics techniques and tools over the next ten years, with the following approaches leading to community-wide adoption:

- stepped roll-out of pilot projects;
- recurring workshops and community engagement activities;
- development of data literacy through education; and
- advocacy at the agency level for nanoinformatics as an essential piece of the nanotechnology research and development enterprise.

2. THE NANOINFORMATICS COMMUNITY

2.1 AN INTEGRATIVE VIEW OF NANOTECHNOLOGY

The nanotechnology community is diverse. Nanotechnology is an inherently integrative field, useful across many commercial sectors and rooted in multiple fields of science and engineering. Real-world applications of nanotechnology have the potential to revolutionize the medical and pharmaceutical industries, address critical issues in the energy and utilities sector, create materials and devices with new functionalities, improve food, paper, electronics and textile products, and make important contributions to national security and defense.

Nanotechnology draws participants from disciplines ranging from materials science, chemistry, and physics to biology, engineering, and environmental health and safety.[6] It is a field that relies heavily on experimentation, computation, simulation, and networked communication. Yet it is also a field where, despite large amounts of data being generated for research in academic and industry laboratories, there is a *lack of reliable, discoverable data* that is standardized, verified, and capable of being shared effectively[7].

This deficiency can be addressed through coordinated development of nanoinformatics tools and methods, allowing community members to validate and leverage data produced by others in ongoing nanotechnology R&D programs. Leveraged, validated data can guide the design of new products, the integration of nanotechnology into large-scale manufacturing, and the analysis of environmental, health, and safety impacts of engineered nanomaterials. One aspect of this strategy for analyzing available information is to provide the highest quality information and data to all stakeholders, including government regulators, industry, and the public. Impacts will include enhanced public understanding and greater success in commercialization. As such, this dimension of nanoinformatics address one of the key goals set forth by the NNI—supporting the responsible development of nanotechnology. To facilitate this goal, effective and open channels of communication between relevant stakeholders must be established.

In their March 2010 assessment of the NNI, the Presidential Council of Advisors on Science and Technology (PCAST) provided one key recommendation impacting nanoinformatics: "Support wide distribution and availability of new non-proprietary information about the properties of nanomaterials."[8] A long term nanoinformatics roadmap must address the challenge of developing mechanisms that allow broad access to EHS-relevant data by state and local regulatory and enforcement agencies as well as the public, while understanding and respecting issues associated with confidential business information that may impact industry. Mechanisms must also be developed for rapidly disseminating information generated within the federal and state governments and agencies. To identify and address nanotechnology EHS issues in a timely manner, the period between information gathering,

---

[6] Porter AL, Youtie J, Shapira P, Shoeneck DJ. 2008. Refining search terms for nanotechnology. Journal of Nanoparticle Research 10(5): 715 – 728. DOI: 10.1007/s11051-007-9266-y

[7] See for example, IRGC White Paper on Nanotechnology Risk Governance (2006).

[8] Report to the President and Congress on the Third Assessment of the National Nanotechnology Initiative, March 12, 2010.

stakeholder input, and information dissemination must be shortened significantly. To address these aspects of information aggregation and sharing, new approaches to database development, analysis, and accessibility must be explored and established. The tools and expertise from the field of information science, which has been actively addressing the issues of data management across domains, will be a helpful guide as nanoinformatics progresses.

2.2 GRASSROOTS

A number of nanoinformatics projects are already underway, and have begun to collect, curate, disseminate, and analyze nanotechnology information. These efforts are supported through a variety of federal, state, and industrial sources. For example, there are ongoing efforts to collect and provide access to difficult-to-obtain information on nanomaterial properties and safe handling data, such as the Nanoparticle Information Library[9], the Nanomaterial – Biological Interactions Database[10], the ICON *Good*Nano *Guide*[11], and the Nanomaterials Registry[12]. InterNano[13] provides information important to the nanomanufacturing community, including process techniques, reports, and taxonomy of terms. The NanoHUB[14] curates computational tools for data analysis, simulation, and education. In addition, there are a handful of cross-institutional efforts to harmonize nanomaterials data and facilitate the interoperability of data projects, such as the caBIG Nanotechnology Working Group[15], or to establish standard terminology metadata within a major domain cluster, such as the Nano Particle Ontology[16] for nanotechnology cancer research.

These existing initiatives have developed independently or *ad hoc* within specific communities of practice, including for example EHS, cancer research, modeling and simulation, and manufacturing. These initiatives emanate from multiple institutions and operate on varying levels of financial and administrative support. While coordination and cross-fertilization among these projects would mitigate redundancy and build on their complementarity, there has been no overarching plan to coordinate these diverse efforts to date. There has been no organized effort to project realistic, community-wide goals for a functional nanotechnology informatics infrastructure that would benefit the nanotechnology community at large, nor is there dedicated funding or agency support for such a comprehensive approach.

2.3 SYNOPSIS OF CURRENT PROJECTS

---

[9] http://nanoparticlelibrary.net/
[10] http://nbi.oregonstate.edu
[11] http://www.goodnanoguide.org/
[12] http://bit.ly/gwEFzd
[13] http://www.internano.org
[14] http://nanohub.org/
[15] http://sites.google.com/site/cabignanowg/home
[16] http://www.nano-ontology.org/

Despite the lack of formal coordination, the nascent efforts in nanoinformatics are making considerable progress independently and have begun the process of reaching out to one another for collaborative advancement. The organizations and projects listed below represent the notable nanoinformatics efforts to date; all have come into being in the last decade.

NANOHUB.ORG

Host: Network for Computational Nanotechnology (NCN)
Domain: Simulation, Education
Established: 2002
Funded by: NSF
URL: http://www.nanohub.org

Established in 2002 by the National Science Foundation, the Network for Computational Nanotechnology (NCN) has a mission to create a national resource for theory, modeling and simulation in nanotechnology, to connect users in research, education, design and manufacturing. That mission is embodied in their web site at nanoHUB.org, which serves more than 140,000 users each year.

CoSMIC

Host: Iowa State University
Domain: Materials; Combinatorial Science
Established: 2002
Funded by: NSF, DARPA, Department of Homeland Security, Army Research Office, Office of Naval Research
URL: http://cosmic.mse.iastate.edu/

CoSMIC is an international collaborative research program focused on data driven discovery in materials science. Its central research theme is to develop new computational and experimental ways of accelerated mechanistic based discovery and design of materials using informatics methods. The nanoinformatics aspect of the program explores how informatics can be used to elucidate nanoscale mechanisms in materials, develop a rational design strategy for new nanomaterials and enhance the quantitative analysis of spectral and imaging data at the nanoscale. Applications of the research include, discovering new nanocluster chemistries of materials, extracting pico-scale information from high resolution imaging and other characterization techniques and integrating nanomaterial data curation with informatics.

NANO EHS VIRTUAL JOURNAL; *GOOD*NANO *GUIDE*

Host: International Council for Nanotechnology (ICON)
Domain: Environmental Health and Safety
Established: 2004; 2009
Funded by: NSF
URL: http://icon.rice.edu/

The International Council on Nanotechnology is a multi-stakeholder organization established in 2004 at Rice University through the Center for Biological and Environmental Nanotechnology. ICON's goal is to develop and communicate information regarding the potential environmental and health risks of nanomaterials to decision makers in industry, government, academia, and civil society. ICON's free resources include a comprehensive news service, the NanoEHS Virtual

Journal—a repository of citations of published research on nanomaterial health, environmental and occupational impacts (http://icon.rice.edu/virtualjournal.cfm)—and the more recent *Good*Nano *Guide*, an online, community-based resource for collecting and sharing information and good practices for safe and responsible handling of nanomaterials. (http://goodnanoguide.org)

NANOPARTICLE INFORMATION LIBRARY

Host: Oregon State University and National Institutes for Occupational Health and Safety (NIOSH)
Domain: Occupational Health
Established: 2004
Funded by: NIOSH
URL: http://nanoparticlelibrary.net

The Nanoparticle Information Library (NIL) was also established in 2004 as part of the NIOSH Nanotechnology Research Program (www.cdc.gov/niosh/topics/nanotech). The NIL is a searchable database of nanomaterials property information and associated health and safety information designed to help occupational health professionals, industrial users, worker groups, and researchers organize and share information on nanomaterials. The NIL provides real-life examples of nanomaterials and their associated origins, properties, and applications and can be used to support the development of a number of needed environmental health and safety tools, training aides, guidelines and standards.

In addition, the NIOSH Field Teams conduct visits to nanomaterial producers and users to characterize exposures, evaluate controls, and develop best practices.

NANOTECHNOLOGY CHARACTERIZATION LABORATORY; CANANOLAB; AND CABIG NANO WORKING GROUP

Host: National Cancer Institute
Domain: Cancer Research; Nanomedicine
Established: 2005; 2009
Funded by: NIH
URL: https://cananolab.nci.nih.gov/

The Nanotechnology Characterization Laboratory (NCL) was established in 2005 to conduct and standardize the characterization of nanomaterials intended for cancer therapeutics (http://ncl.cancer.gov/). The NCL's charter is to serve as a national characterization facility for nanomaterials submitted from academia, industry, and other government laboratories. Among the NCL's objectives is the goal of establishing and standardizing analytical cascade for nanomaterial characterization for use by multiple stakeholders.

caNanoLab is an information portal to facilitate data sharing and standards development within the cancer research community. It builds upon the National Cancer Institute's cancer BioInformatics Grid (caBIG), an established component of the infrastructure supporting cancer research and translation to the clinic (https://cabig.nci.nih.gov/). caNanoLab has been developed in collaboration with the NCI Center for Biomedical Informatics and Information Technology, the NCL, and the NCI Cancer Centers of Nanotechnology Excellence.

In addition, the National Cancer Institute (NCI) Cancer Biomedical Informatics Grid (caBIG®) established a Nanotechnology Working Group in 2009 for researchers with a specific interest in informatics and computational approaches to nanotechnology, with a particular emphasis on nanomedicine. The goal of this working group is to demonstrate the scientific potential of federating nanotechnology databases through pilot projects aimed at integrated semantic search and retrieval of nanomedicine and nanotoxicology data sets that are applicable across nanoscience. The caBIG® Nanotechnology Working Group (caBIG® Nano WG) comprises over 20 active participants from academia, government and industry with diverse interests.

WINGS

Host: Luna Innovations
Domain: Environmental Health and Safety; Occupational Health
Established: 2007
Funded by: US Air Force
URL: unavailable

WINGS, Web Interface Nanotechnology Guidance System, is an Air Force-funded, centralized online resource for nanotech ESOH information from trusted sources, organized into a searchable database and information network. WINGS uses present guidance framework based on industry- and academia-forged nanotechnology EHS frameworks developed at Luna and UDRI. WINGS is primarily a method by which the Air Force personnel can obtain the latest and most accurate information on nanotechnology ESOH information, but it also facilitates information sharing among branches of the military and other government agencies, academic institutions, and so forth, for nanotechnology ESOH information.

INTERNANO

Host: National Nanomanufacturing Network (NNN)
Domain: Nanomanufacturing
Established: 2008
Funded by: NSF
URL: http://www.internano.org

InterNano

The National Nanomanufacturing Network, established through the Center for Hierarchical Manufacturing, launched its information service, InterNano, in 2008. InterNano supports the information needs of the nanomanufacturing community by bringing together resources about the advances in processes, applications, devices, metrology, and materials that will facilitate the commercial development and/or marketable application of nanotechnology. InterNano includes custom nanomanufacturing process and organization databases and utilizes a unique taxonomy to enhance information discoverability.

ACTION-GRID

ACTION-Grid[17] is a European Commission-funded support action, coordinated by the Universidad Politécnica de Madrid (Spain). Since June 2008, this initiative has aimed to establish links between

17 http://www.action-grid.eu

Biomedical Informatics, Grid computing and the novel field of Nanoinformatics. In this regard, it has been the first initiative in this area at the European level, introducing the nano or atomic dimension into the Virtual Physiological Human[18] framework by supporting the establishment of the basis and foundations for nanoinformatics. It is also aimed at expanding international cooperation in the field. Partners and experts from Europe, the Western Balkans, Africa, Latin America and the USA participated in the project, preparing a White Paper[19] suggesting a Roadmap with recommendations and priorities for Nanoinformatics.

Host: Universidad Politécnica de Madrid
Domain: Biomedical Informatics, Nanomedicine
Established: 2008
Funded by: European Commission
URL: http://www.action-grid.eu/

ACTION-Grid finished on May 2010, but members of the project are still addressing scientific and engineering issues related to Nanoinformatics, such as: data and knowledge bases of nanoparticles and biological interactions; creation of nano-ontologies; research on interoperability and integration; data mining from large databases; representations and models for nanoparticles; and connection of databases of toxic effects with computerized medical records.

NANOMATERIAL-BIOLOGICAL INTERACTIONS KNOWLEDGEBASE

Host: Oregon State University
Domain: Biology; Environmental Health and Safety
Established: 2009
Funded by: ONAMI, NSF, NIH, EPA, AFRL
URL: http://nbi.oregonstate.edu/

The Nanomaterial-Biological Interactions Knowledgebase is intended to offer industry, academia and regulatory agencies a mechanism to rationally inquire for unbiased interpretation of nanomaterial exposure effects in biological systems. The unique physicochemical properties of nanomaterials and the inherent complexity of biological systems dictate the immediate need for such an expert system for conceptualizing, analyzing, and visualizing data on fundamental nanomaterial-biological interactions. NBI was designed to enhance dissemination of critical data and information on nanomaterial hazards to industry, academia, regulatory agencies and the general public. This expert system is being further developed to predict the toxic potential of unsynthesized nanomaterials, provide the computational and analytic tools to suggest material design or redesign that may minimize hazard, and propose experimental platforms/methods most predictive of nanomaterial-biological interactions. Features of the NBI knowledgebase will allow for unbiased interpretations of nanoparticle-biological interactions, discovery of unique structural

18 http://www.vph-noe.eu/background
19 http://www.action-grid.eu/index.php?url=whitepapernano

characteristics that govern nanomaterial-biology interactions, and determination of critical data required to predict effects from nanomaterial exposure.

NANOMATERIAL REGISTRY

Host: Research Triangle Institute, International
Domain: Nanomaterials
Established: 2010
Funded by: NIBIB, NIEHS, NCI
URL: pending

The Nanomaterial Registry is a recently-awarded project to establish an authoritative web-based nanomaterial registry that will contain and make available curated information on the biological and environmental interactions of well-characterized nanomaterials; that will be interoperable with existing resources; and that will employ a web-based portal to support efficient data searching, querying, and reporting.

2.4 CONVERGENCE

By year-end 2010, these pioneering nanoinformatics initiatives have generated a critical mass of data, tools, and expertise through their unique projects. Cooperative development now seems obvious. Coordinating efforts even at the most fundamental level is a necessary progression for the work already completed and will facilitate work to come. Such coordination will be of service to the entire spectrum of nanotechnology research and development efforts.

Also by year end 2010, there has been well-documented concern within the scholarly community at large over the stewardship of federally funded research data.[20] The importance of managing, maintaining, and accessing data is seen not only as a mechanism to expedite the scientific process but also as an expectation for the ethical conduct of research, and it is rapidly becoming mandatory practice for many funding agencies. The National Science Foundation, to provide one recent example, now requires all grant proposals to include a two-page data management plan.[21] Scholarly organizations, such as SPARC (Scholarly Publishing and Academic Resources Coalition)[22] and the Association of Research Libraries[23], have been integrating resources on data management into their corpus to aid those who provide direct support to researchers in the field.

Further, the President's Council of Advisors on Science and Technology (PCAST) issued a report in December 2010 on the importance of ongoing federal investment in Networking and Information Technology (NIT). Among other things, the report calls for NIT investments in support of national priorities, specifically energy and transportation, sensors, and high-performance computing—all areas

[20] Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Prepared for the National Academies by the Committee on Science, Engineering, and Public Policy 2009.
[21] http://www.nsf.gov/bfa/dias/policy/dmp.jsp
22 http://www.arl.org/sparc/index.shtml
23 http://www.arl.org/

where nanotechnology can be effectively applied. Further, the report calls for coordination and support for effective cyberinfrastructure that will support new research areas and paradigms.[24]

Additionally, trade publications are covering the topic of informatics and data management with increasing frequency. A recent article published by IndustryWeek illustrates the importance of integrated, secure data management practices as drivers of innovation for industry.

> "In a climate where time-to-market timelines are continually shrinking, the research and innovation side of the house must be more closely aligned with the development and manufacturing side. This where an end-to-end web services-based foundation for information sharing and collaboration comes into play….Organizations need to be able to easily access and integrate critical data across the entire product design and development pipeline so that issues such as environmental fate and safety can be factored into the product lifecycle from day one."[25]

Insofar as nanoinformatics efforts are directed toward expediting nanotechnology research and development and pushing information as well as products into the public arena, it is as much a concern for industry as it is for academic and governmental interests.

It bodes well for the future of nanotechnology R&D that the community's initial efforts are beginning to mature just as these national trends toward large-scale infrastructure and systematic curation come to the fore. Major nanotechnology research and development programs are entering their critical second decade at a time when enhanced use of valid and meaningful data is increasingly recognized as an essential contributor to American competitiveness in scientific innovation.[26] Now is an ideal time to identify and coordinate complementary efforts at the intersection of nanotechnology and informatics, with the common goals of accelerating progress in nanotechnology research and facilitating rapid discovery and innovation.

3. NANOINFORMATICS 2010: A COLLABORATIVE ROADMAPPING WORKSHOP

3.1 OVERVIEW

Nanoinformatics 2010 followed two small-scale, precursor events in nanoinformatics that were hosted by the National Nanomanufacturing Network (NNN). The first was an exploratory nanoinformatics meeting held in June of 2007[27] to identify and prioritize nanoinformatics needs, discuss ongoing

---

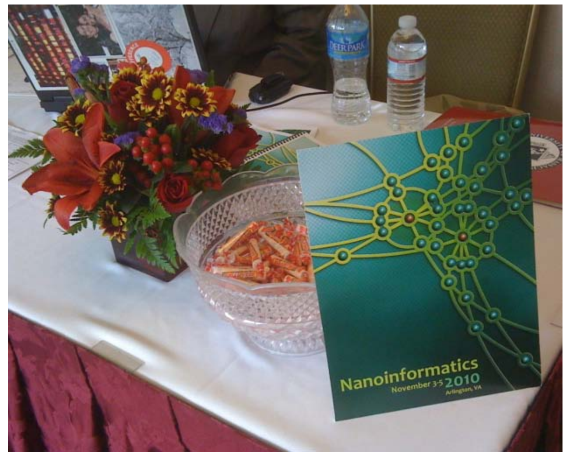
[24] Designing a Digital Future: Federally-funded Research and Development in Networking and Information Technology. December 2010.

[25] Doyle M. *From Science to Sustainability*. [Internet.] IndustryWeek August 11, 2010.

[26] US Nanotech Leadership Faces Global Challengers [internet]. Small Times August 19, 2010; Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. (2007)

[27] http://128.119.56.118/~nnn01/NewFiles/Nanoinformatics_Agenda.htm

activities, and outline strategies for the future. This workshop brought together computer scientists, research scientists, engineers, and others engaged with nanoscale science and engineering as well as with informatics technologies in disciplines such as medicine or materials research. Two years later, in conjunction with its May 2009 Nanomanufacturing Summit[28], the NNN hosted a panel on "Information Needs for Nanomanufacturing" to showcase initiatives geared toward the collection and exchange of critical information for the nanomanufacturing community. Participants in the 2009 event issued a clear call for coordination and roadmapping among diverse nanoinformatics initiatives and expressed strong support for a major, dedicated nanoinformatics event.



On this recommendation, Nanoinformatics 2010 was designed to survey the current nanoinformatics landscape and to stimulate collaborative activities and pilot projects. Nanoinformatics 2010 was planned around the belief that cooperation between disciplines and organizations and the advancement of nanotechnology research and development can be facilitated through informatics efforts. Accordingly, organizers promoted the event widely and invited the participation of all stakeholders concerned with streamlining the nanotechnology data lifecycle and addressing the problem of sparse data in the multidimensional field[29] of nanomaterials property information.

Over 70 workshop participants from 45 national and international organizations attended the two-and-a-half day event from November 3 – 5, 2010, in Arlington, Virginia. Workshop participants included research scientists from academic and government laboratories; informatics specialists; government and industry representatives; and leaders from existing nanoinformatics projects (See Appendix 3).

The Nanoinformatics 2010 Workshop was organized around three distinct but overlapping themes:

- Data Curation and Collection;
- Tools for Innovation, Analysis, and Simulation; and
- Data Accessibility and Information Sharing.

The program itself was designed to bring together diverse disciplinary-based approaches to the workshop themes; to showcase nanoinformatics tools and projects; to discuss challenges associated

[28] http://www.internano.org/ocs/index.php/NMS/NMS2009
[29] For more on the idea of a multidimensional space, See Baker NA, et.al. 2009. Nanotechnology Informatics White Paper. Prepared for the National Cancer Institute by the caBIG® Integrative Cancer Research Nanotechnology Working Group, pp. 10.

with each theme; and to enable the community to sketch out a path for addressing the grand challenges of nanotechnology through e-Science.

Featuring project demonstrations, topical presentations, posters, and discussion, the workshop demonstrated a critical mass for forward progress in nanoinformatics across disciplines and engendered a synergistic atmosphere conducive to networking and the creation of new ideas for interaction among participants. In addition, the presentations, breakout sessions, and discussions articulated areas for development within each theme, as well as illustrated cross-cutting issues that transcend the thematic architecture of the workshop and present as the key challenges that must be addressed before significant, community-wide adoption of efficient practice can be achieved.

3.2 CROSS-CUTTING ISSUES

Out of the many discussions and themed presentations that took place during Nanoinformatics 2010, some foundational elements emerged that determine the direction that new and existing projects need to take to move nanoinformatics forward effectively.

NANOINFORMATICS IS EMERGING AS A VITAL PART OF THE RESEARCH AND DEVELOPMENT TOOLKIT.

- There is vast untapped value in the growing base of data produced by the nanotechnology research community.
- Accessible data—and suitable tools to mine, visualize, model and utilize it—will accelerate the beneficial impacts of nanotechnology, manage risks, and increase efficiency and cost effectiveness.

Limited resources are a reality. Information gaps are always present. Activities in nanoinformatics provide added value in filling information gaps and create pathways for data-driven discovery and innovation. The last 10 years in nanotechnology research have produced a wealth of data. Unlike genome data, the data of the broad fields of nanotechnology is diverse and inhomogeneous. Deeper, more complete data sets are needed, especially to support the utilization of several specific examples of high-potential nanomaterials. Automation and high-throughput screening methods should be developed and utilized. Commensurate tools for mining, modeling, simulating, visualizing, and applying nanotechnology data promise great impact, but are only in their infancy of development.

To realize the largest impact from nanotechnology, more data should be accessible. A modest amount already exists in open databases, but more is desired. Open access data makes web-based utilization tools feasible. Some data will be accessible, but not through the web. Other data will be accessible only through proprietary channels. Even though some data sets may not be openly accessible, it is still very useful to know of their existence, for example through metadata archived in open institutional or disciplinary repositories.

Focused projects that target specific high-interest scientific or technical objectives will demonstrate the intrinsic value of nanoinformatics data and tools. Each concentrated effort will identify information gaps and spur the development of informatics tools that address the specific needs of a compelling research and development topic. Champions have strong incentive to participate in such projects when the objectives are sufficiently specific to impact their own professional concerns. At the same time through coordination and education efforts, the practices, methods and data developed in the course of such activities can be disseminated and leveraged across the entire field of nanoinformatics.

COORDINATION AND INCENTIVE ARE ESSENTIAL FOR SUCCESSFUL NANOINFORMATICS.

- Tools and pilots should leverage existing databases and technologies for application to specific scientific problems.
- The semantic web, open and linked data, and open science initiatives offer informatics technologies and practices that can be leveraged to speed nanotechnology research and development.

Problems, projects, and tools currently exist; nanoinformatics initiatives ought to build upon progress that has already been made and find ways to leverage the existing expertise, data, and infrastructure for discrete scientific inquiries before creating new tools or databases. For example, there are several projects focused on creating and analyzing nanomaterial data. Nanomaterial data is foundational to many cross-domain investigations; as new data sets are produced, for maximum impact, they should utilize minimal data standards to enable facile search and federation capabilities. Coordination and comprehensiveness among existing and future data collection projects must be pursued strategically and supported.

One avenue for facilitating the coordination of projects is through the mechanisms offered by the Semantic Web. The Semantic Web, conceived broadly to include the initiatives of Open Data, Linked Data, and Open Notebook Science, is a set of practices and standards designed to enable individuals to structure their data in ways that are compatible with web-based exchange. The Semantic Web "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."[30]

STANDARDS FOR DATA DOCUMENTATION ARE CRITICAL.

- Standardization of and minimum requirements for data/information will facilitate sharing and evaluation of data sets.
- Industry will be an important participant on standards activities for nanotechnology.

[30] http://www.w3.org/2001/sw/

Establishing standards for minimum requirements for data content and for data quality will be a critical first step toward facilitating effective exchange of information among and between research projects. While this might be driven by the agency-supported programs that have already demonstrated activity in the areas of ontology and controlled vocabularies, it is critical that industry be involved in any deliberations over minimum standards for data sets. Interactions with documentary standards organizations, including ISO, ASTM, IEEE and others, will help to strengthen the effectiveness and impact of nanoinformatics data standards.

SUCCESSFUL INFORMATICS IS A TECHNO-SOCIAL ISSUE.

- The value of nanoinformatics must be well articulated and demonstrated in each of the various fields: health, materials, manufacturing, etc.
- Coordinated advocacy via consortia will represent the value/importance of nanoinformatics projects to the National Nanotechnology Coordination Office and to individual funding agencies.
- Guidelines for sharing and proof-of-concept projects will help to ameliorate any cultural reticence toward information sharing, provided intellectual property interests are adequately represented.

Nanoinformatics projects must be seen to have value to all areas of application; or, projects need to be carried out in multiple areas of application to effect a community-wide adoption of appropriate techniques and practices for data sharing. While EHS is already playing a lead role, the value of nanoinformatics must also be made evident to others in the scientific and commercial value chain so that participation in a community of sharing is seen as an accelerator rather than a barrier to advancement. Communication and educational components should be an inherent part of each distinct nanoinformatics project, so that the major features and capabilities of each tool and data set can be easily understood and utilized across the nanoinformatics community.

In addition, advocacy for nanoinformatics ought to be engaged at the federal and state levels to encourage support of projects that utilize nanoinformatics techniques, standards, and best practices. Bottom-up demonstrations of value together with top-down advocacy for nanoinformatics can enable the community at large to embrace and utilize nanoinformatics in routine workflows.

Furthermore, establishing a successful nanoinformatics program will require demonstration of nanoinformatics' efficacy with respect to intellectual property rights for researchers and companies. It is one thing to show proof-of-concept from a technical vantage point; it is something else to show proof-of-concept that appeals to the financial and legal concerns of various stakeholders. Addressing the issues of intellectual property rights, confidential business information, attribution, and citation are first steps to enable proper documentation of data for exchange.

3.3 WORKSHOP THEMES

The following is a summary of the themed breakout discussions, describing the issues and challenges faced for each. Discussion centers on identifying the barriers to nanoinformatics activity, with recommended steps and solutions to address these challenges offered in section 4.2 as pilot projects.

THEME 1: DATA COLLECTION AND CURATION

This theme was concerned with how researchers in both academia and industry obtain and manage their own data, as well as how they discover data generated by others. This theme encompasses topics such as database management, instrumentation, high-throughput data collection, nanometrology, and semantic compatibility between disciplines.

1. Minimal information standards for nano-data sets (completeness and quality) are required.

Discussions centered on the key information that would be required to distinctly identify a nanomaterial and the assay platform, particularly relevant to the context of nanomedicine or EHS. No attempt was made to define critical physico-chemical characterization parameters due to the fact that numerous other groups (MINChar Initiative, OECD Working Party on Manufactured Nanomaterials, ISO Technical Committee 229) have had exhaustive debate about minimal characterization needs. It is widely accepted that what will be considered the minimal information requirements for one field, may be insufficient for another. For instance, if “size” were a suggested parameter of interest for nano-data sets, that characteristic may need to be additionally refined for application to structure-activity relationships (SARs). To define nanomaterial-specific SARs, information on the primary particle size, size distribution, state of agglomeration, etc. may be required information to sufficiently describe the nanomaterial. Information for modeling would have potentially different/additional requirements.

Environmental, health and safety (biocompatibility) was focused on heavily because it was thought to be a common denominator for all industry and research sectors since safety evaluations are critical to the commercialization of new nano-enabled products. Group discussions focused on the need to disseminate the information requirements that will enable data sharing. The importance of understanding how the data was processed was a key point made by several parties. Participants identified a critical need to have a mechanism/process (tool) whereby the completeness and quality of a data set could be evaluated. With regard to quality, for example, materials need to be characterized using scientifically accepted practices and techniques, as well as some calibration standard for the equipment and tools utilized. The discussion of the “minimal information required” largely transitioned to a sliding scale for describing attributes of a particular data set, allowing potential users of the information to have access to a quality factor associated with the information when choosing a particular data set. Such an evaluation tool could be set by individual data-repositories and/or end users of the data in order to meet their preferred requirements. For example, developers of the Nanomaterial Registry would be able to set their data standards for completeness and quality very high since they are positioned to be an authoritative resource for nanomaterial information. However, researchers at academic institutions may find value in data that may not be complete, but is of high quality. Further discussions included retaining raw data from characterization as an additional dimension of the database for users that may get more useful information from that type of data.

2. Inter-laboratory studies (ILS) to support data are required.

Defining the structural characteristics and physico-chemical properties that biologically active nanomaterials possess is essential to identifying the features that are predictive of biological and environmental responses and consequentially, the material modifications that can minimize hazard. However, the current lack of data on the biological activity, environmental impact and toxic potential of nanomaterials limits *a priori* predictions of environmental safety and biocompatibility. Given the breadth of complementary research efforts aimed at bettering our understanding of nanomaterial-biological interactions, coordination of activities is immediately necessary to ensure functional integration and sharing of data/information, to improve the efficiency of information transfer from data to knowledge, and to reduce the incidence of duplicative studies. A massive research/data coordination network would enhance global participation and support an international research paradigm that overcomes our current limitations to identifying the fundamental principles that govern nanomaterial exposure and nanomaterial-biological interactions. A Nanotechnology Research Coordination Network would 1) facilitate the sharing of engineered nanomaterials and data/information on their physico-chemical properties, environmental transformations/interactions and biological activity, 2) coordinate interdisciplinary, collaborative research efforts focused on investigating nanomaterial-biological interactions, 3) communicate networking efforts and educational outreach opportunities to the international scientific community and 4) provide expertise to government, academia and industry on the biological activity and toxic potential of nanomaterials.

3. Standardized characterization is needed community-wide.

Discussions centered on identifying the common data elements required to sufficiently identify a nanomaterial. With the acknowledgement that many groups have gone through this exercise, the goal would be to leverage work done by the other groups to define a minimal information standard for data sharing (for identifying a nanomaterial as unique). It was suggested that the minimal set of data elements for describing a nanomaterial not be altogether burdensome so as to lower the threshold for "compliance." Size was used as an example of the ideal type of information that we would want to have (size, size distribution, instrument used for evaluating size) and the data that would be essential to have across all studies/materials evaluations, which may just be size. Efforts are currently limited by the disparity in reporting nanomaterial characterization, including the minimal information to describe inherent nanomaterial properties (e.g., size, shape, composition, surface chemistry, surface area, purity, solubility) as well as interactive properties of the nanomaterials (e.g., biological responses at the molecular, cellular, and whole organism level; environmental fate and transport; uptake, bioavailability, biomagnification) which are dependent on exposure/study scenarios (e.g., exposure media and route, duration and timing of exposure, dose/concentration).

Discussions further addressed the type of data or information that would be archived in a database, including raw data from the actual characterization tests, which would have different values to different users accessing information from the database. Suggestions to address this included having different levels of information stored within the material data record that could be accessed, but didn't clutter the top level, core information archived for the material, which would entail the "minimal" data set.

4. How much information is needed to trigger a "recognized hazard"?

Discussion of this issue included what minimal set of information on a material was sufficient to trigger a recognized hazard or safety concern. The discussion considered whether this would be tied into the minimal data set requirements and additionally whether a data quality factor could be tied in with this aspect. Typical requirements for type and for how much data is needed for a given material include some level of toxicity studies complimenting size and composition information as an example. The question remained as to whether such information would be capable of predicting hazards from materials sets based on similarities in the data prior to full-blown toxicity studies.

THEME 2: TOOLS AND METHODS FOR DATA INNOVATION, ANALYSIS, AND SIMULATION

This theme was concerned with the development and use of nanoinformatics tools. Topics in this theme address workflow, data mining and semantic search, gap analysis tools, modeling and simulation, machine learning, and visual analytics.

1. A complete map of workflow guides the development of nanoinformatics.

In any nanotechnology workflow—be it for fundamental research, product development, manufacturing, or another work domain—critical information is needed to support the chain of work activities from inception to completion. A clear map of workflow helps to utilize informatics data and tools to its greatest extent: to optimize outcomes, to increase efficiency, to realize new opportunities. In a developing field such as nanotechnology, and the specific topical activities within, there are gaps and deficiencies that prevent efficient workflow execution. At present, many of these gaps are informational (e.g., insufficient robust data and missing predictive tools). In some cases it is not clear which data is missing, or which information could catalyze new progress. Discovery gap analysis—steered by a collaboration of the interested stakeholders and utilizing informatics software tools—can help to identify critical gaps and topics for focused activity. It must be underscored that the data development activity should be intimately guided by the needs of each specific community desiring that data. For time and cost effectiveness, the production of new desirable data should utilize the state-of-the-art in laboratory automation, data dissemination, information analysis and efficient collaboration techniques. Herein lay great opportunities for nanoinformatics to accelerate workflow in all areas of nanotechnology.

2. A mechanism for federated searches is needed to utilize existing nanotech databases.

A growing set of nanotechnology databases and information repositories exist, but without an effective mechanism to search them all. Each repository has been developed in an *ad hoc* manner, each with its own design features. Some are open access whereas others are not. Some have APIs, whereas others do not. Although it is recognized that within a longer-term framework, standards should be developed enabling database federation with efficient search capabilities, in the short-term a metacrawler should be developed that takes advantage of the databases in their present, heterogeneous form. If an enterprise-like approach is pursued, this would eventually result in a "Nanogrid" designed to connect data and tools. The initial phase would involve discovery of existing data and tools, description of inputs and outputs, APIs, and the initial grid design. This would be followed by the creation of an operational pilot and development of interoperability features. A top-down versus bottom-up balance between system-wide design-standards, rigidity, and local database design flexibility is sought.

3. Getting the science right.

The value and utility of nanoinformatics must be clearly demonstrated. Although broadly applicable, the inherent value is best demonstrated on a case-by-case basis. Modeling and simulation have already proven their worth in various areas of science and industry, but there are many other areas in which such activities have not yet been developed and deployed. It is important that computational nanoscience tools get the science right. To do so, they must be subject to validation and verification with reference nanoparticles and nanomaterials. New codes must be verified with well-understood test sets and test problems. A standard reference system to compare and validate codes would help to facilitate such activities. The emphasis should shift away from the developer of the tool to the accurate performance of the tool. As such, collaborative code development is advisable. Virtual vaults of code components, for example pseudopotentials, can provide greater opportunity for getting the science right. A simulation challenge, targeting specific materials and involving a blind prediction challenge, is advisable. Sensitivity analysis should become an inherent activity in all modeling, simulation and design work.

4. Getting the right data.

Materials development is driven by specific scientific and industrial need. Since there are thousands of various nanomaterials and hundreds of properties to be measured, it is important to target a smaller set of specific high-use or high-potential materials and the most relevant properties. Whereas open availability of data is desired, it is recognized that this is not always possible. Nonetheless, in cases where data is not openly accessible it is important to disseminate broadly that such data exists and the pathways (e.g., pricing or permissions) through which it can be accessed. There are many data gaps that exist in current scientific, industrial, and safety workflows in nanotechnology. In addition to physical properties, the statistical distribution around an average value of each relevant property is needed for sensitivity analysis, which is crucial for predictability, design, manufacturing, and safety. Coalitions of strong champions in specific subfields can be a useful mechanism to identify data needs and to develop a plan to gather the right data. Several of the current nanoinformatics projects do exactly that, for example the caBIG Nanotechnology Working Group for nanomedicine. However, other areas are "data sparse" with no clearly identified plan for obtaining the data. Combined, physical data, modeling, and simulation tools create the opportunity for the development of new predictive nanoinformatics workflow tools. One way that this could be realized is to create data-driven simulation/optimization challenges in specific sub-fields of nanotechnology where a significant amount of data and theory already exists, but significant gaps are present as well. Pilots in such strategic areas could make significant progress and identify specific targets for intensely focused research and development.

5. Tools, training, and education.

To support a federated system of nanoinformatics tools, it is necessary that each resource provide metadata that describes each tool so that the broader community can easily discover it. In addition, basic educational and training features should supplement each tool so that users can quickly understand its applicability and use. Providing an XML document of I/O of each tool is one example. Bridges with commercial software tools can strengthen the interaction with the nanoinformatics R&D community. A broader range of visualization, analysis, and design tools are needed for nanotechnology. One example is structural model visualization. The creation of visualization challenges on important nanomaterials or systems could catalyze progress in innovative ways. This is both a research and educational issue, driven by both fundamental research and industrial needs.

THEME 3: DATA ACCESSIBILITY AND INFORMATION SHARING

This theme was concerned with the practical, cultural, legal, and ethical aspects of sharing data within the nano community. Such issues include proper data annotation and attribution (citation analysis, DOIs), the cultural dynamics of data sharing in various disciplines, barriers to data sharing, governance and regulation of data and intellectual property, and issues of data interoperability and federation, metadata, and standards development.

1. Context is critical for effective information sharing.

The ability to document and communicate the context of a given data set will facilitate information sharing, particularly across domains. The notion of context correlates well to the idea of provenance for scientific workflows. Provenance provides information about the processes used to create a data product, information that enables the data to be reproduced and clarifies its dependencies, determines the data's authorship and quality, and assists in preserving the data.[31] Going further, robust contextual information would lend itself to a broader understanding of the role that the data plays in a scientific inquiry and might determine how the data can best be applied and reused. For example, nanomaterial

---

[31] Davidson and Friere. 2008. Provenance and Scientific Workflows: Challenges and Opportunities. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. DOI: 10.1145/1376616.1376772

characterization data derived from a research study on high-density storage may include valuable but incomplete data for a related materials toxicity study. Ready availability of the provenance information indicating the genesis, purpose, and area of application for a given data set would dictate the potential uses of a data set, identify the degree to which a given data set might be incorporated into new studies, drive risk analysis, and help to align perception with reality. Context parameters should be part of any minimum requirements for data sharing.

2. Competing socio-cultural incentives impact data sharing and must be addressed.

The range of expectations for or against data sharing that a given investigator is subject to will impact their ability or interest in fully participating in a culture of data sharing. Therefore it is important not only to identify what that range of expectations may be, but also to offer a tiered approach to participation that would allow investigators to share their data without compromising their obligations to collaborators, employers, and funders.

A variety of scenarios might prevent an investigator from sharing their information with the broader community, such as:

- The traditional publishing paradigm discourages open sharing of unpublished results;
- Disciplinary standards do not exist;
- Due to the nature of the research, it is either inadvisable or inappropriate for data to be shared;
- Data sets contain Confidential Business Information (CBI); or
- Funders require confidentiality.

At the same time, factors that might compel, or in some cases require, an investigator to share their information are beginning to become more common, such as:

- Disciplinary practices encourage the open sharing of data;
- Open sharing facilitates cross-institutional collaboration;
- The magnitude of the domain itself calls for open systems to expedite progress (i.e., astronomy, climate change); or
- Funders mandate the sharing of primary research data.

This range of competing cultural influences for and against data sharing presents an opportunity to identify a spectrum of classifications for sharing and to offer investigators a variety of options for attribution and/or collaboration through DOIs or other best practices.

To implement any such system effectively, it is useful to identify classes of issues that impact data sharing and to evaluate their level of importance for different stakeholder groups, as in Table 1. This exercise begins to illustrate those areas where solutions must be in place before information sharing can be conducted effectively and with respect to an investigators' and/or funders' intellectual property rights.

A clear outlier here is the ability to annotate, attribute, and integrate data with services. Understanding the diversity of needs on this issue across stakeholder groups would enable a solution to this issue to be either identified or articulated and would incentivize data sharing. A considerable barrier to attribution is the fact that large society and STEM publishers exert significant influence against pre-publication sharing of information. By enabling consistent, persistent identifiers for data sets, as is done with large-scale data repositories and projects such as DataCite[32], investigators would be able to share their data in a citable, discoverable manner that would not interfere with standard professional procedures, such as publication and tenure. Looking to communities that have already devised alternative solutions to these problems—such as the nanoHUB.org and the Open Notebook Science initiative—would offer models and best practices to follow going forward.

	Policy makers and regulators	Researchers (Academic/Industry/Volunteers)	IT/informaticists	Management	Legal community	Equipment and system vendors	Health and safety practitioners	Educators	Funders	End users	Media	Society	Workers
Annotation, Attribution, and Author Services	x	A/V/I	x		x	x	x		x				
Interoperability and federation/ Metadata	x	A/I	x			x				x			
Aligning perception with reality	x			x						x	x	x	
IP, Security, Open Access/Open Data		A/V/I	x	x	x				x				
Communicating science		A		x				x			x	x	
Standards Development	x	A	x			x							
Governance and Regulation	x			x	x		x						

Table 1: Issues that need to be addressed to implement effective information sharing for specific stakeholder groups, as identified during the Theme 3 breakout discussion.

32 http://datacite.org/

Insofar as the use of metadata standards is a matter of awareness and practice, cultural dynamics are important for the adoption of standards by community members. To encourage the use of standards throughout an interdisciplinary domain such as nanotechnology, both traditionally conservative and forward-thinking disciplines engaged in nanotechnology research and development must be willing to embrace the tools and mechanisms that will enable progress.

4. A ROADMAP FOR NANOINFORMATICS

4.1 A DECADE-LONG VISION

This Roadmap is proposing a decade of growth, beginning with initial activities focused primarily around workshops and pilot projects. The Roadmap is intended to be dynamic, being updated periodically. Workshops review recent progress in nanoinformatics and identify critical needs and emerging opportunities. Pilots mobilize fast action on a small number of priority topics. The workshops, pilots, and broader nanoinformatics R & D activity all feed the continued development of the Roadmap. The Nanoinformatics 2010 Workshop demonstrated current projects, opened a discussion of potential solutions to nanoinformatics problems and drivers, and established new pilot projects to solve those problems. The workshop will be followed by a period of work by distinct groups on specific problems—the pilot projects—that will come together in a follow up workshop in late 2011 to report on progress and update the nanoinformatics roadmap for 2012.

The first few years of activity are truly critical as they will demonstrate the willingness of the nanoinformatics community's commitment to moving forward. The Roadmap will identify cross-cutting issues that impact the long-term vibrancy of nanoinformatics and propose a path forward. The pilot projects outlined are foundational and will serve to produce more activity in following years.

Through year five, foundational projects and advocacy will make nanoinformatics an essential component of the nanotechnology research and development enterprise. It is further expected that during this time additional areas of pilot development will be established as new themes evolve through the ongoing workshops and discussion amongst participants. Critical to these emerging themes will be both input and adaptation by industry to better address the key needs of this essential portion of nanotechnology stakeholders.

Moving toward the ten-year perspective, the Roadmap will identify objectives that move toward a robust, collaborative nanoinformatics system that will have a demonstrated impact on the scientific and societal aspects of nanotechnology.

4.2 PILOT PROJECTS

Pilot projects are intended to demonstrate the feasibility and showcase the impact of nanoinformatics on specific, tightly-focused topics. Cooperative efforts can demonstrate successful implementation of projects with low investment and significant results, while laying the groundwork for later, more extensive efforts.

To the maximum extent possible, these pilots will leverage cooperative activities already funded and underway. In some cases, new resources will be needed to realize pilot activity. In all cases, each proposed pilot is not intended to duplicate a similar effort that may be of interest to a particular funding agency. Rather, it is the intent that each proposed pilot be adopted or absorbed by one or more funding source, and that, through affiliation with this Nanoinformatics Roadmap community, each pilot is provided with more comprehensive set of expert resources and a platform from which to achieve and showcase progress.

Eleven pilots proposed during the workshop have been consolidated into seven complementary, one-year pilot projects, described below. Two of them are concerned with engagement; two are focused on metadata and standards; and three are geared toward tools development and deployment. As "one-year" pilots, they are expected to make some definitive progress within the first year of activity. This does not preclude ongoing work or future activities. The "one-year" designation is a helpful mechanism to spur focused activity within each group.

ENGAGEMENT PILOTS

1. Consortium for Coordinating Nanomaterials Research Data

Given the breadth of complementary nanomaterials research efforts underway, coordination of activities is necessary to ensure functional integration and sharing of data/information, to improve the efficiency of information transfer from data to knowledge, and to reduce the incidence of duplicative studies. Currently, there is no program or agency liaison to coordinate between various organizations and enforce the standardization of data set information.

This pilot seeks to establish a consortium, for example a Nanotechnology Research Coordination Network, for coordinating between the various organizations for such activities as issuing data set quality factors, establishing ILS calibrations, and ensuring that the necessary requirements and information exists for follow-on risk assessment studies. In addition, such a consortium would coordinate interdisciplinary, collaborative research efforts; communicate networking efforts and educational outreach opportunities; and provide expertise to government, academia, and industry on nanomaterials.

The outcome of this pilot would be, minimally, a proposal for funding to establish a dedicated nanomaterials informatics consortium. Impact would be community-wide and foundational, potentially reaching all sectors engaged with nanomaterials for study or commercial application.

2. Workshops for Focused Nanomaterials Development Using Nanoinformatics

This pilot will run workshops to target two specific nanomaterials of high potential impact to use as scientific drivers and areas of proof-of-concept assessment for the application of nanoinformatics methods. The two topics will focus on a specific area within the field of nanocomposites and a specific area of nanomedicine. Topics are chosen which already have a substantial base of literature and data, and for which some informatics tools already exist. These will start with virtual workshops and follow up with an in-person meeting including participants from the industrial and research sectors.

Each workshop will frame the outcome in terms of materials challenge, for example, materials by design and use of the web for materials development. It is a priority to work with industry and trade organizations. The workshops will determine type of data and information needed with emphasis on physico-chemical properties, EHS and other desired data. The workshop will engage suppliers and users of nanomaterials, modelers, experimentalists, and informatics specialists. The objective is to use nanoinformatics tools to identify scientific information gaps and inform funding agencies of high priority topics as potential areas for support.

METADATA AND STANDARDS PILOTS

3. Meta-ontology for Cross-discipline, Cross-sector Information Exchange

The diversity of domain-specific ontologies and taxonomies for nanotechnology R&D is an impediment to broad-based and effective information and knowledge sharing. Creating upper-level ontology and demonstrating its applicability across multiple domains would provide a common vocabulary for the nanotechnology community and present a facile mechanism for the sharing of data among complementary but distinct research programs.

The Meta-ontology Pilot will focus on integrating and rationalizing standards already in use, and defining interactions between concepts as validated and reusable methods that deliver value to the stakeholders. This pilot is designed to model existing knowledge in a way that it can be correctly delivered. The two activities are clearly complimentary and in concert would deliver more value to the user. The first activity will be addressed by creating an abstract core ontology that would eventually allow stakeholder groups to map their taxonomies and semantic web ontologies to the core, although initially that would be done by the project team. Multiple and overlapping terms can coexist and be managed by contextual relevance and equivalence maps. The second activity will be addressed by defining each concept in the core ontology as a set of formalized quantitative and/or qualitative scenarios, including (but not limited to) rules, formulas, fuzzy logic, standards, measures, methods for validation and sensitivity, as well as contextual parameters. As a result, rather than aiming for a static definition of each concept, additional scenarios can be proposed as science and technology evolves and vetted scenarios can be added to the core. The evolving model can become an engine for some aspects of validation and development of new theories.

Primary output of this pilot would be a core nano ontology. Two of our critical success indicators would be 1) usefulness and usability across a range of stakeholders that is likely to grow, and 2) scientific validity or verifiability that is consistent across diverse stakeholder groups, which could include regulatory agencies, discovery researchers, and product engineers spanning multiple application domains. All Nanotechnology R&D stakeholders (e.g., researchers, product designers, and regulators from diverse application areas) would potentially be impacted by such an overarching approach.

4. Minimum Information Requirements for Data Sharing (Completeness and Quality)

Present materials characterization provides data sets that may include a range of analytical characterization techniques providing specific properties for a given nanomaterial being studied. This information is then made accessible through a given platform database where the information is archived. As databases are further developed, or a given data set is expanded upon, no specification of information requirements are provided, either in terms of data quality or completeness.

This pilot will determine the minimal information required for nanomaterials data sets, both in terms of completeness and quality. Activities will determine the necessary information requirements for data sets to enable sharing and/or incorporating data within pre-existing databases in such a way that a quality factor can be associated with each data set and that all data sets contributed to a given database have some standards for further sharing and use. This includes specifications of analytical techniques used to characterize a given nanomaterial, and further provides the basis for standardization of these techniques, along with how the data is actually processed and archived.

The outcome of the pilot would be a list of standard materials information and characterization techniques, as well as the minimum data set necessary to obtain a specified quality factor to be included within a database. Any researcher desiring to access and use the data sets archived within a database have some assurance regarding the types of analysis, characterization, and integrity of the information. This would potentially impact the entire community utilizing nanomaterials.

TOOLS PILOTS

5. Meta-crawler for Mining Nanotechnology Repositories and Open Access Sources

Conventional search engines crawl the Web broadly, not deeply. To obtain all the information possible from each existing nanotechnology/nanomaterial database on a selected material and conduct gap analysis on this collection requires a tool that can deeply and intelligently explore the known nanotechnology databases.

This pilot will create and implement a custom metacrawler to mine the known nanotechnology databases as well as open access nanotechnology resources. Such a project could be used with the goal answering a specific research question and could demonstrate data gaps and help to articulate subsequent calls for action.

In addition, this project will include the objective to recommend a guideline of minimum suggested content for literature abstracts to facilitate metacrawler search and discovery. Although research abstracts and author keywords are required metadata for publication, there is a lack of uniformity for such information among various STEM and society publishers, making the search, retrieval, and mining of such data a challenge. A meta-crawler with requirements for abstract content would not only generate richer and more meaningful search results but also facilitate the systematic mining of such literature for large-scale exploration of literature sets.

6. nano-SAR Education and Dissemination

This project will demonstrate the ability to combine structural data and modeling to develop nanoscale structure-activity relationships (nano-SARs) that can be disseminated as an educational tool; efforts of this pilot will coordinate with materials activities in other pilots and existing projects.

Initial work will focus on assessing and consolidating existing knowledge of nano-SARs from the current literature. Also, the group will define the minimal standards required for performing valid nano-SARs as compared to more general minimal information standards required for the characterization of nanomaterials.

Subsequent work for this pilot will be to create an educational model for structural data that clearly illustrates nanoscale structure-activity relationships. This model will be widely disseminated via the nanoHUB to engender consistent understanding of what is a fundamental element for simulating nanomaterial-biological interactions.

The outcome of this pilot will be a nanoHUB module on nano-SARs with supporting reference and pedagogical material. This pilot's most immediate impact would be among the nanotechnology education community, but could have broader impact as a "textbook" equivalent for structure activity relationships.

7. Simulation Resources and Simulation Challenge

The medical community has highly accurate reference calculations, to help comparisons. Validation and verification is integral to the accuracy of such research tools. For software and model development, a simulation challenge—such as a blind prediction challenge—targeting the properties of a specific nanomaterial could produce such standard reference systems to compare and validate data emerging from such calculational tools. Such a challenge would need to have well-defined goals, identify data gaps, and include mechanisms of sensitivity analysis.

For simulation and modeling, a target material must be selected. The choice of material depends on a pressing end goal, a specific property, or combination of properties. For toxicity studies, both a toxic material and nontoxic analogs are needed. A nanomedicine target is another interesting choice. These pilot activities could be coupled as a satellite to an allied conference. The limitations of the physical aspects of the models also need to be considered. Materials that undergo large structural changes, or particles with surfaces that change over time in response to their environment, may exceed current calculational capabilities. Standard nanostructures that are well characterized should be used to compare and validate calculation tools. (Analogous to the use of G20 for Gaussian.)

The desired inputs and outputs need to be addressed. The data and the tools used to associate that data need to be clearly defined. Also, additional property measurements should be identified that are needed to make the challenge robust. The desired tool types—first principle simulation, empirical models, data analysis, data visualization, and data exchange tools—should be clearly indicated, as driven by research or development needs.

4.3 COMMUNICATION AND ASSESSMENT RECOMMENDATIONS

As we move forward with The Nanoinformatics 2020 Roadmap, and with respect to the geographically-distributed and diverse nature of the community, we recommend that existing initiatives and pilot projects build education and communication components into their day-to-day work that utilize networked communication tools in efficient ways and enable productive knowledge transfer among community members. For example,

- Employ metadata standards, such as the NanoParticle Ontology, in current work projects and integrate them into routine workflows and project documentation and procedures;
- Create APIs useful for federation and deployment of web services for nanotechnology research and development;
- Use analytics as metrics for evaluating the impact of data and web services;
- Engage in workshops and bring new mechanisms for nanotechnology research and development to the community where it can be tested and used;
- Use existing tools for dissemination and sharing of data and information;
- Consider using author addenda when submitting manuscripts to closed-access journals, publishing in Open Access journals, or self-archiving manuscripts in institutional or domain appropriate repositories so that research becomes more widely discoverable.

These are just a few examples of ways that day-to-day research activities, either in academic, government, or industry labs, that can make data more readily available to the nanoinformatics community.

APPENDIX 1: GLOSSARY

These definitions of terms used in the Nanoinformatics 2020 Roadmap have been excerpted from various Web resources as indicated. Attribution to the respective resource is given in each case.

API (application programming interface)

An Application Programming Interface (API) is a language and message format used by an application program to communicate with the operating system or some other control program such as a database management system (DBMS) or communications protocol. APIs are implemented by writing function calls in the program, which provide the linkage to the required subroutine for execution. Thus, an API implies that some program module is available in the computer to perform the operation or that it must be linked into the existing program to perform the tasks.

(http://www.pcmag.com/encyclopedia_term/0,2542,t=application+programming+interface&i=37856,00.asp)

Assay

The determination of the amount of a particular constituent of a mixture or of the biological or pharmacological potency of a drug.

(http://www.biology-online.org/dictionary/Assay)

Attribution

Attribution in copyright law is the requirement to acknowledge or credit the author of a work which is used or appears in another work. Attribution is required by most copyright and copyleft licenses, such as the GNU Free Documentation License and Creative Commons licenses.

(http://en.wikipedia.org/wiki/Attribution_%28copyright%29)
(See also: http://creativecommons.org/)

Database

Any collection of data or information that is specially organized for rapid search and retrieval by a computer. Databases are structured to facilitate the storage, retrieval, modification, and deletion of data in conjunction with various data-processing operations.

(http://www.britannica.com/EBchecked/topic/152195/database)

DOI (digital object identifier)

A digital object identifier (DOI) is a name assigned to any entity for use on digital networks. DOIs are used to provide current information, including where the entities (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change. The DOI System provides a framework for persistent identification, managing intellectual content, managing metadata, linking customers with content suppliers, facilitating electronic commerce, and enabling automated management of media.

http://datacite.org/whatisdoi.html

EHS (environmental health and safety)

While the novel behaviors of nanomaterials have the potential to bring about technological advances in many areas, such as energy, medicine, and the environment, these behaviors may also pose a risk to human health and the environment. Nanotechnology environmental health and safety (EHS) is concerned the range of issues associated with evaluating and predicting the potential human and environmental risks posed by nanomaterials and with managing those risks.

(http://www.nano.gov/html/society/EHS.html)

Federated Databases

A federated database system is a type of meta-database management system (DBMS) which transparently integrates multiple autonomous database systems into a single federated database. The constituent databases are interconnected via a computer network, and may be geographically decentralized. Since the constituent database systems remain autonomous, a federated database system is a contrastable alternative to the (sometimes daunting) task of merging together several disparate databases.

(http://en.wikipedia.org/wiki/Federated_database_system)

Federated Search

Federated search is an information retrieval technology that allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user.

(http://en.wikipedia.org/wiki/Federated_search)

Linked Data

Using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods; A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.

http://linkeddata.org/; (http://en.wikipedia.org/wiki/Linked_Data)

Metadata

Literally, "data about data." Structured information describing information resources/objects for a variety of purposes.

(http://www.abc-clio.com/ODLIS/odlis_m.aspx)

Nanoinformatics

Nanoinformatics is the science and practice of determining which information is relevant to the nanoscale science and engineering community, while developing and implementing effective mechanisms for collecting, validating, storing, sharing, analyzing, and applying that information. Nanoinformatics also involves the utilization of networked communication tools to launch and support efficient communities of practice. Nanoinformatics is necessary for intelligent development and comparative characterization of nanomaterials, for design and use of optimized nanodevices and nanosystems, and for development of advanced instrumentation and manufacturing processes. Nanoinformatics also fosters efficient scientific discovery and learning through data mining and machine learning techniques.

http://nanotechinformatics.org/nanoinformatics/index.php/Main_Page)

Nanomanufacturing

Nanomanufacturing is the controllable manipulation of materials structures, components, devices, and systems at the nanoscale (1 to 100 nanometers) in one, two, and three dimensions for large-scale reproducibility of value-added components and devices.

(http://www.internano.org/content/view/200/224/)

Nanomaterials

"Nanomaterials" is a term that includes all nanosized (1 – 100 nm) materials, including engineered nanoparticles, incidental nanoparticles and other nano-objects, like those that exist in nature. When particles are purposefully manufactured with nanoscale dimensions, we call them engineered nanoparticles. A nanostructured material has internal structure that is within the 1 to100 nanometer (nm) range, while the pieces of material themselves are larger than 100 nm.

(http://www.nano.gov/html/facts/faqs.html)

Ontology

An ontology is a formal, explicit representation of knowledge belonging to a subject area: the knowledge is encoded and represented as a hierarchy of concepts (terms / classes) that are described using attributes (e.g., metadata such as preferred name, definition, synonyms, etc.), related using associative relations, and formalized using logical axioms in a machine-interpretable language (e.g., Ontology Web Language or OWL).

(http://www.nano-ontology.org/documentation/frequently-asked-questions#TOC-What-is-an-ontology-)

Provenance

Provenance of data is the record of its origin and the history of any modifications or validations. It is an essential component in the specification and utilization of trusted workflows, enabling result reproducibility, data sharing, and knowledge re-use.

(http://portal.acm.org/citation.cfm?id=1376772)

QSAR (quantitative structure-activity relationship)

Quantitative structure-activity relationship (QSAR) is the process by which chemical structure is quantitatively correlated with a well- defined process, such as biological activity or chemical reactivity.

(http://en.wikipedia.org/wiki/QSAR)

SAR (structure-activity relationship)

The Structure-Activity Relationship (SAR) is a means by which the effect of a drug or toxic chemical on an animal, plant or the environment can be related to its molecular structure. This type of relationship may be assessed by considering a series of molecules and making gradual changes to them, noting the effect upon their biological activity of each change. Alternatively, it may be possible to assess a large body of toxicity data using intelligent tools such as neural networks to try to establish a relationship.

(http://msds.chem.ox.ac.uk/glossary/structure_activity_relationship.html)

Semantic Web

Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

(http://www.w3.org/standards/semanticweb/)

Taxonomy

Taxonomy provides a controlled vocabulary for metadata attributes and specifies relationships between terms in the controlled vocabulary, from simple relationships to more specific and complex ones. In terms of Web sites and portals, a site's taxonomy is the way it organizes data into categories and subcategories.

(http://www.taxonomywarehouse.com/include_resources.asp)

Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible, and understandable assessments; and communicate assessment effectively for action.

(http://www.infovis-wiki.net/index.php/Visual_Analytics)

Workflow

A scientific workflow is a formal specification of a scientific process, which represents, streamlines, and automates the analytical and computational steps that a scientist needs to go through from data set selection and integration, computation and analysis, to final data product presentation and visualization.

(http://www.cs.wayne.edu/~shiyong/swf/swf2010.html)

APPENDIX 2: NANOINFORMATICS 2010 WORKSHOP PROGRAM

Wednesday, November 3

Nanoinformatics Landscape

Goal: To achieve a broad understanding of nanoinformatics activities both through demonstrations of existing nanoinformatics projects and through presentations of informatics activities from the nanotechnology research & development domain as well as exemplar disciplines.

8:00 – 5:15	Registration and Exhibits	
8:00 – 8:30	Breakfast	
8:30 – 8:45	**Welcome** *Travis Earles (Office of Science & Technology Policy)* **Purpose of Workshop** *Mark Tuominen (National Nanomanufacturing Network)*	**Ballroom**
8:45 – 10:05	**Project Demonstrations** *Introduction: Mark Tuominen (National Nanomanufacturing Network)*	**Ballroom**
8:45 – 9:05	**caNanoLab** *Sharon Gaheen (SAIC)*	
9:05 – 9:25	**nanoTAB** *Stacey Harper (Oregon State University)*	
9:25 – 9:45	**Development of a Multi-Criteria Decision Analysis Tool to Support Selection of Nanomaterial Studies** *Gretchen Bruce (Intertox, Inc.)*	
9:45 – 10:05	**Web Interfaced Nanotechnology ESOH Guidance System (WINGS)—An Overview** *Aaron Small (Luna Innovations Incorporated)*	
10:05 – 10:30	Break	
10:30 – 12:00	**Opening Keynotes** *Introduction: Jeff Morse (National Nanomanufacturing Network)*	**Ballroom**
10:30 – 11:15	**nanoHUB.org and the Delivery of Value to Authors and Users** *George Adams (Network for Computational Nanotechnology)*	
11:15 – 12:00	**Nanotoxicology as a Predictive Science That Can Be Explored by High Content Screening and the Use of Computer-assisted Hazard Ranking** *Andre Nel (University of California, Los Angeles)*	
12:00 – 1:00	Lunch	**The National Diner**
1:00 – 2:15	**Theme 1 Presentations: Data Collection and Curation** *Co-Chairs: Stacey Harper, Yoram Cohen, Peter Schad*	**Ballroom**
1:05 – 1:25	**Collaboration and Data Management** *Yoram Cohen (University of California, Los Angeles)*	
1:25 – 1:50	**Nanoparticle Ontology for Cancer Nanomedicine Research** *Nathan Baker (Pacific Northwest National Laboratory)*	
1:50 – 2:15	**The Collaboratory for Structural Nanobiology: Nanoparticle Structural Analysis as a Rosetta**	

	Stone *Raul Cachau (SAIC-Frederick NCL)*	
2:15 – 2:30	Break	
2:30 – 3:45	**Theme 2 Presentations: Tools for Innovation, Analysis, and Simulation** *Co-Chairs: Anne Chaka, Mark Tuominen, Michael McLennan*	**Ballroom**
2:35 – 2:55	**Nanoparticles Toxicity: Knowledge Extraction from High-Throughput Screening Data** *Rong Liu (University of California, Los Angeles)*	
2:55 – 3:20	**Cloud Computing for Science** *Kate Keahey (Argonne National Laboratory)*	
3:20 – 3:45	**Scientific Workflow Tools** *Daniel Crawl (San Diego Supercomputer Center)*	
3:45 – 4:00	Break	
4:00 – 5:15	**Theme 3 Presentations: Data Accessibility and Information Sharing** *Co-Chairs: Michele Ostraat, Mark Hoover, Rebecca Reznik-Zellen*	**Ballroom**
4:05 – 4:25	**Considerations in the Application of Nanoinformatics to Occupational Safety and Health.** *Paul Schulte (NIOSH)*	
4:25 – 4:50	**PhenX Measures for Data Sharing, Cross-study Analysis and Data Interoperability** *Carol Hamilton (RTI International)*	
4:50 – 5:15	**The Implications of Open Notebook Science and Other New Forms of Scientific Communication for Nanoinformatics** *Jean-Claude Bradley (Drexel University)*	
6:00 – 7:00	**Poster Session and Reception**	**The National Diner**
7:00 – 8:00	**Keynote Banquet** *Introduction: Mark Hoover (NIOSH)*	**Ballroom**
7:30 – 7:40	**Remarks from the NanoBusiness Alliance** *Vincent Caprio (NanoBusiness Alliance)*	
7:40 – 8:00	**Remarks on the State of Nanotechnology** *Mihail C. Roco (National Science Foundation)*	

Thursday, November 4

Nanoinformatics Roadmapping

Goal: To stimulate discussion and launch the roadmapping activities that are the primary objective of the workshop, wherein themed groups will focus on a single set of informatics issues in depth and craft specific recommendations to address them.

8:00 – 5:15	Registration and Exhibits	
8:00 – 8:30	Breakfast	
8:30 – 10:00	**Roadmapping Set Up: Additional Perspectives** *Introduction: Vicki Colvin (Rice University)*	**Ballroom**
8:35 – 9:00	**EPA Perspectives on Nanoinformatics for Prioritization and Toxicity Testing** *Sumit Gangwal (National Center for Computational Toxicology)*	
9:00 – 9:25	**Nanoinformatics in Europe: The ACTION Grid White Paper** *Victor Maojo (Universidad Politécnica de Madrid)*	

9:25 – 9:50	**Getting to 'The 5 stars of Linked Open Data' for Nanoinformatics** *Mills Davis (Project10x)*	
10:00 – 10:30	Break	
10:30 – 12:00	**Roadmapping Set Up: Project Vignettes** *Introduction: Phil Lippel*	**Ballroom**
10:35 – 11:35	**nanoHUB.org** *Michael McLennan (Purdue University)* **CoSMIC** *Krishna Rajan (Iowa State University)* **NIST** *Anne Chaka (NIST)* **ICON GoodNanoGuide** *Vicki Colvin (Rice University)* **Nanoparticle Information Library** *Mark Hoover (NIOSH)* **NIOSH Field Teams** *Chuck Geraci (NIOSH)* **NNN/InterNano** *Jeff Morse (National Nanomanufacturing Network)* **caNano Working Group** *Nathan Baker(Pacific Northwest National Laboratory)* **NCL Metadata Projects** *Marty Fritts (Nanotechnology Characterization Laboratory)* **NBI Knowledgebase** *Stacey Harper (Oregon State University)* **Materials Registry** *Michele Ostraat (RTI International)*	
11:35 – 12:00	**Panel Discussion** *Phil Lippel, moderator*	
12:00 – 1:00	Lunch	**The National Diner**
1:00 – 2:15	**Themed Breakouts** *Theme1: Data Collection and Curation* *Theme 2: Tools for Innovation, Analysis, and Simulation* *Theme 3: Data Accessibility and Information Sharing*	 **Ballroom I & II** **Ballroom III & IV** **Eisenhower Room**
Theme 2 additional presentations	**Nanoinformatics from the Biomedical Informatics Perspective** *G.H. López-Campos (Institute of Health "Carlos III")* **Developing a Virtual Vault for Pseudopotentials: A NNIN/C Initiative** *Derek Stewart (Cornell University)* **Using Open-Source Scripting Languages for Rapid-development of Informatics Capabilities** *Craig Versek (University of Massachusetts Amherst)*	
2:15 – 2:30	Break	
2:30 – 3:45	**Themed Breakouts** *Theme 1: Data Collection and Curation* *Theme 2: Tools for Innovation, Analysis, and Simulation* *Theme 3: Data Accessibility and Information Sharing*	 **Ballroom I & II** **Ballroom III & IV** **Eisenhower Room**

3:45 – 4:00	Break	
4:00 – 5:15	**Themed Breakouts**	
	Theme 1: Data Collection and Curation	**Ballroom I & II**
	Theme 2: Tools for Innovation, Analysis, and Simulation	**Ballroom III & IV**
	Theme 3: Data Accessibility and Information Sharing	**Eisenhower Room**
5:15 – 6:00	Break	
6:00 – 7:00	**Reception**	**The National Diner**
7:00 – 8:00	**Networking Dinner**	**Ballroom**

Friday, November 5

Nanoinformatics Roadmapping

Goal: To wrap up the roadmapping activities through report-ins and general discussion and to coordinate the pilot projects which will move activities forward to 2011.

8:00 – 5:15	Registration and Exhibits	
8:00 – 8:30	Breakfast	
8:30 – 10:00	**Report In and General Discussion**	**Ballroom**
	Introduction: Krishna Rajan (Iowa State University)	
8:35 – 10:00	*Anne Chaka (National Institute of Standards and Technology)*	
10:00 – 10:30	Break	
10:30 – 12:00	**Pilot Planning for Nanoinformatics 2011**	**Ballroom**
	Introduction: James Luo (National Institutes of Health)	
10:30 – 12:00	*Marty Fritts (Nanotechnology Characterization Laboratory)*	
12:00 – 1:00	**Closing Keynote Lunch**	**Ballroom**
	Introduction: Mark Tuominen (National Nanomanufacturing Network)	
12:30 – 1:00	*Sylvia Spengler (National Science Foundation)*	

APPENDIX 3: NANOINFORMATICS 2010 PARTICIPANTS

Nanoinformatics 2010 Participant List

FIRST NAME	LAST NAME	AFFILIATION	EMAIL ADDRESS
George Bunch	Adams	NCN	gba@purdue.edu
Carolina	Avendano	Rice University	cavendano@rice.edu
Nathan	Baker	PNNL	nathan.baker@pnl.gov
Amy	Bednar	Army Corps of Engineers	amy.e.bednar@usace.army.mil
Jean-Claude	Bradley	Drexel University	Jeanclaude.bradley@gmail.com
Gretchen	Bruce	Intertox	gbruce@intertox.com
Raul	Cachau	SAIC-Frederick NCL	raul.cachau@ymail.com
Vincent	Caprio	NanoBusiness Alliance	vincentcaprio@nynanobusiness.org
Anne	Chaka	NIST	anne.chaka@nist.gov
Yoram	Cohen	UCLA CEIN	yoram@ucla.edu
Daniel	Crawl	SDSC	crawl@sdsc.edu
Chelsea	D'Angona	OSTP	dangona@gwmail.gwu.edu
Mills	Davis	Project 10x	mdavis@project10x.com
Diana	De la Iglesia	Universidad Politecnica de Madrid	diglesia@infomed.dia.fi.upm.es
David	Dix	National Center for Computational Toxicology	dix.david@epa.gov
Travis	Earles	OSTP	Travis_M._Earles@ostp.eop.gov
Heather	Evans	OSTP	hevans@ostp.eop.gov
Marty	Fritts	NIH/NCI	frittsmj@mail.nih.gov
Sharon	Gaheen	SAIC	sharon.m.gaheen@saic.com
Sumit	Gangwal	EPA	gangwal.sumit@epa.gov
Charlie	Gause	Luna Innovations	GauseC@lunainnovations.com
Chuck	Geraci	NIOSH	cgeraci@cdc.gov
Joe	Glick	Expertool	joeglick@expertool.com
Maneesha	Gupta	Aligarh Muslim University	maneesha.nano@gmail.com
Carol	Hamilton	RTI International	chamilton@rti.org
Stacey	Harper	OSU	Stacey.Harper@oregonstate.edu
Taimur	Hassan	UCLA CEIN	thassan@cnsi.ucla.edu
Liesl	Heeter	NNCO	lheeter@nnco.nano.gov
Mark	Hoover	NIOSH	zij3@cdc.gov
Ali	Hosseini	University of Mazabdaran	hos-a-p1@umz.ac.ir
Matthew S.	Hull	NanoSafe Inc.	mhull@nanosafeinc.com
Kate	Keahey	Argonne National Laboratory	keahey@mcs.anl.gov
Fred	Klaessig	Pennsylvania Bio Nano Systems	fred.klaessig@verizon.net
Chang Sun	Kong	Iowa State University	cskong@iastate.edu
Sharon	Ku	NIH	sharon.ku@nih.gov
Kristin	Kulinowski	Rice University	kk@rice.edu
Pragya	Kushwaha	Panjab University	pragya189@gmail.com
Jeff	Lewandowski	IOP	jeff.lewandowski@iop.org
Philip H.	Lippel		phlippel@gmail.com

Rong	Liu	UCLA CEIN	rongliu@ucla.edu
Victoria	Lopez-Alonso	Carlos III	victorialopez@isciii.es
Guillermo	Lopez-Campos	Carlos III	glopez@isciii.es
James	Luo	NIH	luoja@mail.nih.gov
Bettye	Maddux	SNNI-ONAMI	bmaddux@uoregon.edu
Trouble	Mandeson	NNN UMass Amherst	mandeson@polysci.umass.edu
Victor	Maojo	ACTION Grid	vmaojo@infomed.dia.fi.upm.es
Stephanie	Matthews	University of Georgia	smathe@uga.edu
Charles	McGovern	INSCX Exchange	info@inscx.com
Michael	McLennan	Purdue University	mmclennan@purdue.edu
Jeff	Morse	NNN Umass Amherst	jdmorse@research.umass.edu
Andre	Nel	UCLA CEIN	anel@mednet.ucla.edu
Brand	Niemann	semanticommunity.net	bniemann@cox.net
Michele	Ostraat	RTI International	mostraat@rti.org
David	Paik	Stanford University	david.paik@stanford.edu
Krishna	Rajan	Iowa State University	krajan@iastate.edu
Robert	Rallo	UCLA CEIN	robert.rallo@ucla.edu
Rebecca	Reznik-Zellen	NNN UMass Amherst	rreznikz@library.umass.edu
Pat	Rizzuto	Daily Environmental Report	prizzuto@bna.com
Mihail	Roco	NSF	mroco@nsf.gov
Peter	Schad	RTI International	pschad@rti.org
Paul	Schulte	NIOSH	PSchulte@cdc.gov
Matt	Sedlak	RJLee Group	msedlak@rjlg.com
Amornpun	Sereemaspun	Chulalongkorn University	Amornpun.S@Chula.ac.th
Aaron	Small	Luna Innovations	smalla@lunainnovations.com
Sylvia	Spengler	NSF	sspengle@nsf.gov
Jeff	Steevens	US Army	Jeffery.A.Steevens@usace.army.mil
Bob	Stevens	NNN UMass Amherst	stevens@library.umass.edu
Derek	Stewart	Cornell University	stewart@cnf.cornell.edu
Mike	Thorn	UMass Amherst	michaelt@physics.umass.edu
Mark	Tuominen	NNN UMass Amherst	tuominen@physics.umass.edu
Craig	Versek	UMass Amherst	cversek@physics.umass.edu
Cyrus	Wadia	OSTP	cnwadia@lbl.gov
Diane	Wetherington	Intertox	diane@intertox.com

APPENDIX 4: PILOT PROJECTS TEMPLATES

CONSORTIUM FOR COORDINATING NANOMATERIALS RESEARCH DATA	
Problem Description (What is the problem to be addressed and why is it important?)	Given the breadth of complementary nanomaterials research efforts underway, coordination of activities is necessary to ensure functional integration and sharing of data/information, to improve the efficiency of information transfer from data to knowledge, and to reduce the incidence of duplicative studies.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	Currently, there is no program or agency liaison to coordinate between various organizations and enforce the standardization of data set information.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	This pilot seeks to establish a consortium, for example a Nanotechnology Research Coordination Network, for coordinating between the various organizations for such activities as issuing data set quality factors, establishing ILS calibrations, and ensuring that the necessary requirements and information exists for follow-on risk assessment studies. In addition, such a consortium would coordinate interdisciplinary, collaborative research efforts; communicate networking efforts and educational outreach opportunities; and provide expertise to government, academia, and industry on nanomaterials.
Expected Impact (What impact would the new approach have and who will be impacted?)	Impact would be community-wide and foundational, potentially reaching all sectors engaged with nanomaterials for study or commercial application.
Participants (Who is, or should be, involved?)	UCLA CEIN, ONAMI, PNNL, SAIC, NCL, Penn Nano Bio, NNN, NIST, UPM, other as solicited.
Budget Requirements (How much money is needed?)	Initial funding requirements low, administrative support for the processing of a proposal. Can be absorbed by existing projects.

Pilot Metrics (What are expected metrics or milestones to indicate success?)	Minimally, a proposal for funding to establish a dedicated nanomaterials informatics consortium.

WORKSHOPS FOR FOCUSED NANOMATERIALS DEVELOPMENT USING NANOINFORMATICS	
Problem Description (What is the problem to be addressed and why is it important?)	Utilize existing data and information in the research literature for comprehensive data mining and gap analysis to advance the development of a specific nanomaterial with respect to specific high-priority properties.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	Very few comprehensive data sets exist that enable this to be done from a single resource. There is rich set of data in the literature for various specific nanomaterials, but this information is a patchwork quilt of information that could be.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	The objective is to run workshops to identify specific nanomaterials topics suitable for proving-out the value nanoinformatics through data mining and gap analysis. Each workshop would focus on one specific material type and identify the target material properties to be examined. Effort will focus on materials that already have a substantial base of literature and data. dentification of relevant data resources, informatics tools, and strategic objectives would be workshop outcomes.
Expected Impact (What impact would the new approach have and who will be impacted?)	Proof-of-concept use of nanoinformatics methodologies and tools to extract new value out of existing data, and identification of information/knowledge gaps to nucleate critical new research activities. Identification of challenges and opportunities in executing mining information across heterogeneous data sources.
Participants (Who is, or should be, involved?)	NIST, NNN, NCI, PNNL, industry, and others as solicited
Budget Requirements (How much money is needed?)	In principle, workshops could be co-funded by NNN and NIST. Follow on work could be funded by NSF, NIH and others.
Pilot Metrics (What are expected metrics or milestones to indicate	Workshop execution with strategic plan as output.

success?)	

META-ONTOLOGY FOR CROSS-DISCIPLINE, CROSS-SECTOR INFORMATION EXCHANGE	
Problem Description (What is the problem to be addressed and why is it important?)	Creating an upper-level ontology and demonstrating its applicability across multiple domains would provide a common vocabulary for the nanotechnology community and present a facile mechanism for the sharing of data among complementary but distinct research programs.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	The diversity of domain-specific ontologies and taxonomies for nanotechnology R&D is an impediment to broad-based and effective information and knowledge sharing.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	The Meta-ontology Pilot will focus on integrating and rationalizing what is already in use, and defining interactions between concepts as validated and reusable methods that deliver value to the stakeholders. The two activities are clearly complimentary and in concert would deliver more value to the user. The first activity will be addressed by creating an abstract core ontology that would eventually allow stakeholder groups to map their taxonomies and semantic web ontologies to the core, although initially that would be done by the project team. Multiple and overlapping terms can coexist and be managed by contextual relevance and equivalence maps. The second activity will be addressed by defining each concept in the core ontology as a set of formalized quantitative and/or qualitative scenarios, including (but not limited to) rules, formulas, fuzzy logic, standards, measures, methods for validation and sensitivity, as well as contextual parameters.
Expected Impact (What impact would the new approach have and who will be impacted?)	All Nanotechnology R&D stakeholders (e.g., researchers, product designers, and regulators from diverse application areas) would potentially be impacted by such an overarching approach.
Participants (Who is, or should be, involved?)	PNNL, NNN, Expertool, ActionGRID, independent consultants
Budget Requirements (How much money is needed?)	Funding for Webmeetings and at least one face-to-face meeting. Other costs administrative, to be absorbed by participants.

Pilot Metrics (What are expected metrics or milestones to indicate success?)	Primary output of this pilot would be a core nano ontology. Two of our critical success indicators would be 1) usefulness and usability across a range of stakeholders that is likely to grow, and 2) scientific validity or verifiability that is consistent across diverse stakeholder groups, which could include regulatory agencies, discovery researchers, and product engineers spanning multiple application domains.

MINIMUM INFORMATION REQUIREMENTS FOR DATA SHARING (COMPLETENESS AND QUALITY)	
Problem Description (What is the problem to be addressed and why is it important?)	Inconsistent data sets and a lack of community standards for data completeness or quality make potential data sharing activities cumbersome.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	Present materials characterization provides data sets that may include a range of analytical characterization techniques providing specific properties for a given nanomaterial being studied. This information is then made accessible through a given platform database where the information is archived. As databases are further developed, or a given data set is expanded upon, no specification of information requirements are provided, either in terms of data quality or completeness.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	This pilot will determine the minimal information required for nanomaterials data sets, both in terms of completeness and quality. Activities will determine the necessary information requirements for data sets to enable sharing and/or incorporating data within pre-existing databases in such a way that a quality factor can be associated with each data set and that all data sets contributed to a given database have some standards for further sharing and use. This includes specifications of analytical techniques used to characterize a given nanomaterial, and further provides the basis for standardization of these techniques, along with how the data is actually processed and archived.
Expected Impact (What impact would the new approach have and who will be impacted?)	This would potentially impact the entire community utilizing nanomaterials.
Participants (Who is, or should	UCLA CEIN, ONAMI, PNNL, NCL, SAIC, InterTox, NIST, NNN, ABCC, MINChar,

be, involved?)	Penn Nano Bio, Nanomaterial Registry, Stanford
Budget Requirements (How much money is needed?)	Funding for Webmeetings and at least one face-to-face meeting. Other costs administrative, to be absorbed by participants.
Pilot Metrics (What are expected metrics or milestones to indicate success?)	The outcome of the pilot would be a list of standard materials information and characterization techniques, as well as the minimum data set necessary to obtain a specified quality factor to be included within a database. Any researcher desiring to access and use the data sets archived within a database have some assurance regarding the types of analysis, characterization, and integrity of the information.

META-CRAWLER FOR MINING NANOTECHNOLOGY REPOSITORIES AND OPEN ACCESS SOURCES	
Problem Description (What is the problem to be addressed and why is it important?)	Conventional search engines crawl the Web broadly, not deeply. To obtain all the information possible from each existing nanotechnology/nanomaterial database on a selected material and conduct gap analysis on this collection requires a tool that can deeply and intelligently explore the known nantechnology databases.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	Such a meta-crawler does not currently exist.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	This pilot will create and implement a custom Metacrawler to mine the known nanotechnology databases as well as open access nanotechnology resources. Such a project could be used with the goal answering a specific research project and could demonstrate data gaps and help to articulate subsequent calls for action. In addition, this project will include the objective to recommend a guideline of minimum suggested content for literature abstracts to facilitate search and discovery. Although research abstracts and author keywords are required metadata for publication, there is a lack of uniformity for such information among various STEM and society publishers, making the search, retrieval, and mining of such data a challenge.

Expected Impact (What impact would the new approach have and who will be impacted?)	A meta-crawler with requirements for abstract content would not only generate richer and more meaningful search results but also facilitate the systematic mining of such literature for large-scale exploration of literature sets. Immediate impact would be to the community whose databases are being mined, with subsequent impact to the broader nanotechnology research and development community.
Participants (Who is, or should be, involved?)	ActionGRID, NIST, Stanford, SAIC, NCL, others as solicited.
Budget Requirements (How much money is needed?)	Small budget and in-kind efforts for initial proof-of-concept development.
Pilot Metrics (What are expected metrics or milestones to indicate success?)	Successful extraction of data from disparate, known nanotechnology resources.

nano-SAR EDUCATION AND DISSEMINATION	
Problem Description (What is the problem to be addressed and why is it important?)	There is a need to demonstrate the ability to combine structural data and modeling to develop structure-activity relationships and disseminate as an educational tool.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	Utilize the nanoHUB to create an educational model for correct modeling of structural data and to illustrate structure-activity relationships. This tool would be widely disseminated via the nanoHUB to engender consistent understanding of what is a fundamental element for simulating nanomaterial-biological interactions. Further, this effort should be coordinated with nanomaterials activities.

Expected Impact (What impact would the new approach have and who will be impacted?)	This pilot's most immediate impact would be among the nano education community, but could have broader impact within the nanotechnology community as a "textbook" equivalent for structure activity relationships.
Participants (Who is, or should be, involved?)	nanoHUB, Collaboratory for Sturctural Nanobiology, NCL, SAIC, Army, PNNL, ORNL, UCLA CEIN
Budget Requirements (How much money is needed?)	
Pilot Metrics (What are expected metrics or milestones to indicate success?)	The outcome of this pilot would be a nanoHUB module on structure activity relationships with supporting tutorial and lecture material.

SIMULATION RESOURCES AND SIMULATION CHALLENGE	
Problem Description (What is the problem to be addressed and why is it important?)	A simulation and modeling challenge targeting the properties of a specific nanomaterial or nanomaterial system to establish reliable materials-by-design, nanomanufacturing-by-design, or structure-property predictability.
Current Practice (How is the problem currently being addressed [if at all], by whom, and what are limitations to current practice?)	Many simulation and modeling activities currently exist for nanomaterials or nanomaterials systems, but the reliability and usefulness of these tools is widely inconsistent.
Proposed New Approach (What is the new approach to solving the problem and why is it time to use this approach now?)	The challenge incentivizes new ideas to result in computational science tools that are amenable to validation and verification. Standard nanostructures that are well characterized would be utilized to compare and validate calculational tools.
Expected Impact (What impact would the new approach have and who will be impacted?)	Begin to build nanoinformatics tools with reliable predictive value. Identify the challenges in doing so.

Participants (Who is, or should be, involved?)	NCN, NNN, broader NSE research community.
Budget Requirements (How much money is needed?)	Substantial funding is required to perform this pilot in a comprehensive fashion. Smaller funding to identify a meaningful challenge.
Pilot Metrics (What are expected metrics or milestones to indicate success?)	Identify a few valuable simulation challenge topics. Identify suitable approaches for each and engage the appropriate research communities and funding agencies.

APPENDIX 5: COMMUNICATION AND ASSESSMENT GUIDELINES

As we engage in the development and application of a useful and robust Nanoinformatics 2020 Roadmap, we are endeavoring to apply the following communication and assessment guidelines that we have adapted from the American Statistical Association (ASA) "Guidelines for assessment and instruction in statistics education (GAISE)" (http://www.amstat.org/education/gaise/):

1. Emphasize literacy and develop critical thinking;
2. Develop and use real-life data examples;
3. Stress conceptual understanding rather than mere application of procedures;
4. Foster continuous improvement and active discussions;
5. Use technology for developing conceptual understanding and for analyzing and sharing information (e.g., modeling and simulation, databases, etc.); and
6. Use assessments to improve and evaluate the efficacy and impact of these activities.

In applying these guidelines, we recognize that there are many stakeholders and community roles that have informatics needs and can make informatics contributions. As illustrated in the following table, the nanoinformatics partners include workers, health and safety practitioners, management, policy makers and regulators, equipment and system vendors, the legal community, the finance community, insurers, researchers, educators, the media, consumers, and society in general. Our goal is to use innovative actions in support of the Nanoinformatics 2020 Roadmap to Engage the community, Inform the interested, Reward the responsive, and Understand and incentivize the reluctant.

A Communication and Education Message and Audience Planning Tool for the Nanoinformatics 2020 Roadmap

	Workers	Health and Safety Practitioners	Management	Policy Makers and Regulators	Equipment and system vendors	Legal community	Finance Community	Insurers	Researchers	Educators	Media	Consumers	Society
Literacy and Critical Thinking													
Real Life Examples													
Understanding (not rote application)													
Continuous Improvement													
Modeling and Sharing													
Assessment													

mark.hoover@cdc.hhs.gov 304-285-6374. The matrix can be used to clarify what information stakeholders need and they can provide.

APPENDIX 6: NANOINFORMATICS BIBLIOGRAPHY

CONFERENCE PROCEEDINGS

Allarakhia M, Walsh S, Wensley A. 2007. Models of cooperation and knowledge management: The case of biomedical technology management. [presentation] PICMET '07: Portland International Center for Management of Engineering and Technology Vols. 1 – 6 Proceedings – Management of Converging Technologies: 438 – 447. Conference of the Portland-International Center for Management of Engineering and Technology (PICMET 2007), Portland OR, August 5 – 9, 2007.

Chau M, Chen H, Qin J, Zhou Y, Qin Y, Sung WK, McDonald D. 2002. Comparison of Two Approaches to Building a Vertical Search Tool: A case study in the nanotechnology domain. [presentation] Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital Libraries. International Conference on Digital Libraries, Portland OR, July 14 – 18, 2002. DOI: 10.1145/544220.544246

Chiesa S, Garcia-Remesal M; de la Calle G, de la Iglesia D, Bankauskaite V, Maojo V. 2008. Building an Index of Nanomedical Resources: An automatic approach based on text mining. [presentation] Knowledge-Based Intelligent Information and Engineering Systems, Pt 2, Proceedings 5178: 50-57. 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Zagreb Croatia, September 3 – 5, 2008.

de la Iglesia D, Chiesa S, Kern J, Maojo V, Martin-Sanchez F, Potamias G, Moustakis V, Mitchell JA. 2009. Nanoinformatics: New challenges for biomedical informatics at the nano level. [presentation] Medical Informatics in a United and Healthy Europe - Proceedings of MIE 2009 – The XXIInd International Congress of the European Federation for Medical Informatics, Vol 150: 987 – 991. The XXIInd International Congress of the European Federation for Medical Informatics, Sarajevo, August 30 – September 2, 2009. DOI: 10.3233/978-1-60750-044-5-987

Harper S. 2010. Resource Identification: Forming the Future Interoperable, Federated System [unpublished] Prepared for the caBIG Nanotechnology Working Group Face-to-Face Meeting, February 3, 2010.

Hey, Tony. 2010. The Fourth Paradigm. [plenary] 31st Annual IATUL Conference, West Lafayette, IN, June 20 – 24, 2010.

Lyons, KW. 2007. Integration, Interoperability, and Information Management: What are the key issues for nanomanufacturing? [presentation] Proceedings of SPIE Vol. 6648: 66480D. SPIE Optics and Photonics, August 29, 2007. DOI:10.1117/12.735615.

Martin-Sanchez F, Lopez-Alonso V, Hermosilla-Gimeo I, Lopez-Campos G. 2008. A primer in knowledge management for nanoinformatics in medicine. [presentation] Knowledge-Based Intelligent Information and Engineering Systems, Pt 2, Proceedings 5178: 66-72. 12th International conference on Knowledge-Based Intelligent Information and Engineering Systems, Zagreb Croatia, September 3 – 5 2008.

Molnar LK. 2007. Nanobioinformatics: The enabling technology of personalized medicine. [presentation] Proceedings of the 7th IEEE International symposium on Bioinformatics and Bioengineering, Vols. 1 – 11: 11. 7th IEEE International Conference on Bioinformatics and Bioengineering, Boston MA, October 14 – 17, 2007.

Nelson B, Friedman D, Geisz J, Albin D, Benner J, and Wang Q. 2003. To data management and beyond….for photovoltaic applications. [presentation] Materials Research Society Symposium Proceedings, Vol 804: 295 – 300. Combinatorial and Artificial Intelligence Methods in Materials Science II, Boston MA, December 1 – 4, 2003.

Nowak S. 2003. Technical Nanosystems Based on the Biological Solutions. [presentation] Third IEEE Conference on Nanotechnology Proceedings Vol II: 195 – 198. Third IEEE Conference on Nanotechnology, San Francisco CA, August 12 – 14, 2003. DOI: 10.1109/NANO.2003.1231749

Paik DS. 2007. Toward a Nanobioinformatics Infrastructure for Nanotechnology-based Prostate Cancer Therapeutic Response Tracking. [presentation] Proceedings of the 7th IEEE International Symposium on Bioinformatics and biogengineering, Vols. 1 and 11: 486. 7th IEEE International Conference on Bioinformatics and Bioengineering, Boston MA, October 14 – 17, 2007.

Postek MT and Lyons K. 2007. Instrumentation, Metrology, and Standards: Key elements for the future of nanomanufacturing. [presentation] Proceeding of SPIE Vol. 6648: 664802. SPIE Optics and Photonics, August 29, 2007. DOI: 10.1117/12.730855.

Reznik-Zellen R, Stevens B, Thorn M, Morse J, Smucker MD, Allan J, Mimno D, McCallum A, and Tuominen MT. 2008. InterNano: e-Science for the nanomanufacturing community. [poster] eScience, pp.382-383, 2008 Fourth IEEE International Conference on eScience, Indianapolis IN, December 7 – 12, 2008. DOI: 10.1109/eScience.2008.142

Ruping K, Sherman BW. 2004. Nanoinformatics: Emerging computational tools in nano-scale research. [presentation] NSTI NanoTech 2004 Vol 3 Technical Proceedings: 525-528. Nanotechnology Conference and Trade Show (Nanotech 2004), Boston MA, March 7 – 11, 2004.

Stokes TH, Phan J, Quo CF, Nie SM, Wang MD. 2006. Bio-nano-informatics: An integrated information management system for personalized oncology. [presentation] 2006 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1 – 15: 6425-6428. 28th Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society, New York NY, August 30 – September 3, 2006. DOI: 10.1109/IEMBS.2006.259752

Stukowski M, Suh C, Rajan K, Tall PD, Beye AC, Ramirez AG, Soboyejo WO, Benson ML Liaw PK. 2006. Informatics for Combinatorial Experiments: Accelerating data interpretation. [presentation] Combinatorial Methods and Informatics in Materials Science 894: 301-306. Symposium on Combinatorial Methods and Informatics in Materials Science held at 2005 MRS Fall Meeting, Boston MA, November 28 – December 1, 2005.

Thomas DG, Pappu RV, Baker NA. 2009. NPO: Ontology for cancer nanotechnology research. Available from Nature Precedings. DOI: 10.1038/npre.2009.3514.1

Thomas DG, Pappu RV, Baker NA. 2009. Ontologies for Cancer Nanotechnology Research. [presentation] Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Society: Engineering the Future of Biomedicine, EMBC 2009, 4158-4161. 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, Minneapolis MN, September 2 – 6, 2009.

ARTICLES

Aourag H, Saidi F, Broderick S, Rajan K. 2009. Designing Superlattices Ultra Hard Coatings: Datamining Approach. Journal of Computational and Theoretical Nanoscience 6(4): 828-833. DOI: 10.1166/jctn.2009.1114

Balachandran PV, Broderick SR, and Rajan K. 2011. Identifying the "inorganic gene" for high temperature piezoelectric perovksites through statistical learning. Proceedings of the Royal Society A (in press)

Bowman DM, Ludlow K. 2009. Filling the Information Void: Using public registries as a tool in nanotechnologies regulation. Bioethical Inquiry 6: 25-36. DOI: 10.1007/s11673-009-9134-9http://dx.doi.org/10.1007/s11673-009-9134-9.

Chau M, Huang Z, Qin J, Zhou Y, Chen H. 2006. Building a scientific knowledge web portal: The NanoPort Experience. Decision Support Systems 42: 1216-1238. DOI: 10.1016/j.dsss.2006.01.004

Davidson and Friere. 2008. Provenance and Scientific Workflows: Challenges and Opportunities. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. DOI: 10.1145/1376616.1376772

De la Iglesia D, Maojo V, Chiesa S, Martin-Sanchez F, Kern J, Potamias G, Crespo J, Garcia-Remesal M, Keuchkerian S, Kulikowski C, Mitchell JA. 2011. International efforts in nanoinformatics research applied to nanomedicine. Methods Inf Med. 50(1):84-95.

Ellenbecker M, and Tsai SJ. 2008. Interim Best Practices for Working with Nanoparticles [documentation]. Center for High-rate Nanomanufacturing. (Unpublished) Available from: http://eprints.internano.org/34/

Gil Y, Deelman E, Ellisman M, Fahringer T, Fox G, Gannon D, Goble C, Livny M, Moreau L, Myers J. 2007. Examining the Challenges of Scientific Workflows. Computer 40(12): 24-32 DOI: 10.1109/MC.2007.421

Gold A. 2007. Cyberinfrastructure, Data, and Libraries Pt. 1. D-Lib Magazine 13 9/10. DOI: 10.1045/september20september-gold-pt1

Klimeck G, McLennan M, Brophy SP, Adams GB, Lundstrom MS. 2008. nanoHUB.org: Advancing Education and Research in Nanotechnology. Computing in Science and Engineering 10(5): 17-23. DOI: 10.1109/MCSE.2008.120

Kulinowski, K. M. and Jaffe MP. 2009. "The GoodNanoGuide: A novel approach for developing good practices for handling engineered nanomaterials in an occupational setting." Nanotechnology Law & Business 6(1): 37-44.

Li X, Petersen L, Broderick S, Narasimhan B, and Rajan K. 2011. Identifying Factors Controlling Protein Release from Combinatorial Biomaterial Libraries via Hybrid Data Mining Methods; ACS Combinatorial Sciences 13: 50-58

Maojo V, Martin-Sanchez F, Kulikowski C, Rodriguez-Patron A, Fritts M. 2010. Nanoinformatics and DNA-Based Computing: Catalyzing Nanomedicine. Pediatric Research 67(5): 481-489.

McKinney BA. 2009. Informatics approaches for identifying biologic relationships in time-series data. Wiley Interdisciplinary Reviews—Nanomedicine and Nanobiotechnology 1(1):60-68. DOI: 10.1002/wnan.012

Miller AL, Hoover MD, Mitchell DM, Stapleton BP. 2007. The Nanoparticle Information Library (NIL): A Prototype for Linking and Sharing Emerging Data. Journal of Occupational and Environmental Hygiene 4(12): D131-D134, 2007.

Ostrowski AD., Martin T, et al. 2009. "Nanotoxicology: characterizing the scientific literature, 2000-2007." Journal of Nanoparticle Research 11(2): 251-257.

Porter AL, Youtie J, Shapira P, Shoeneck DJ. 2008. Refining search terms for nanotechnology. Journal of Nanoparticle Research 10(5): 715 – 728. DOI: 10.1007/s11051-007-9266-y

Porter, AL, Youtie J. 2009. How interdisciplinary is nanotechnology? Journal of Nanoparticle Research 11: 1023-1041. DOI: 10.1007/s11051-009-9607-0

Rafols I. 2007. Strategies for Knowledge Acquisition in Bionanotechnology: Why are interdisciplinary practices less widespread than expected? Innovation 20(4): 395-412. DOI: 10.1080/13511610701760770

Schulte P, Geraci C, Zumwalde R, Hoover M, Castranova V, Kuempel V, Murashov V, Vainio H, and Savolainen K. 2018. Sharpening the focus on occupational safety and health of nanotechnology. Scandinavian Journal of Work, Environment & Health 34(6): 471-478. Available from: http://www.sjweh.fi/show_abstract.php?abstract_id=1292

Whitling, JM, Spreitzer G, Wright DW. 2000. Combinatorial and informatics approach to CdS nanoclusters. Advanced Materials 12(18): 1377-1380. DOI: 10.1002/1521-4095(200009)12:18<1377::AID-ADMA1377>3.0.CO;2-X

Wilkins-Diehr N, Gannon D, Klimeck G, Oster S, Pamidighantam S. 2008. TeraGrid Science Gateways and Their Impact on Science. Computer 41(11): 32-41. DOI: 10.1109/MC.2008.470

BOOKS AND BOOK CHAPTERS

Hey T, Tansley S, and Tolle K eds. 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research. 284 pp. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Hofmann DWM and Kuleshova LN (eds). 2010. *Data Mining and Inorganic Crystallography* in Data Mining in Crystallography ; Series: Structure and Bonding , Vol. 134 pp.59-89; Springer-Verlag

C. Kamath C, Wade N, Karypis G, Pandey G, Kumar V, Rajan K, Samatova NF, Breimyer P, Kora G, Pan C, and Yoginath S. 2009. *Scientific Data Analysis* in Scientific Data Management. Shoshani A and Rotem D (eds). Taylor and Francis pp.263-301

Walsh S and Medley T. 2008. A Framework for Responsible Nanotechnology [chapter 18]. Fisher E, et al. (eds.). The Yearbook of Nanotechnology in Society, Vol. 1. Springer Science+Business Media B.V. 2008. pp. 207-213. DOI: 10.1007/978-1-4020-8416-4_18

Rajan K. 2008. Combinatorial Materials Sciences: Experimental Strategies for Accelerated Knowledge Discovery; Annual Reviews of Materials Research 38: 299-322. DOI: 10.1146/annurev.matsci.38.060407.130217

WHITEPAPERS

Baker NA, Fritts M, Guccione S, Paik DS, Pappu RV, Patri A, Rubin D, Shaw SY, Thomas DG. 2009. Nanotechnology Informatics White Paper. Prepared for the National Cancer Institute by the caBIG® Integrative Cancer Research Nanotechnology Working Group.

Baker NA, Fritts M, Guccione S, Paik DS, Pappu RV, Patri A, Rubin D, Shaw SY, Thomas DG. 2009. Ontologies in Cancer Nanotechnology Research. Prepared for the National Cancer Institute by the caBIG Integrative Cancer Research Nanotechnology Working Group.

Baker NA, Freund ET, Gaheen S, Guccione S, Paik DS, Pappu RV, Patri A, Rubin D, Shaw SY, Thomas DG. 2009. The Need for Minimum Information Standards for Development and Advancement of Nanomaterials as Cancer Diagnostics and Therapeutics. Prepared for the National Cancer Institute by the caBIG Integrative Cancer Research Nanotechnology Working Group.

IRGC White Paper on Nanotechnology Risk Governance (2006). Available from: http://www.irgc.org/IMG/pdf/IRGC_white_paper_2_PDF_final_version-2.pdf

Maojo V, Martín-Sánchez F et al. The ACTION-Grid White Paper on Nanoinformatics (2010). Available from: http://www.action-grid.eu/index.php?url=whitepapernano. Last access: February 2011.

RELEVANT REPORTS

Atkins D, et. al. 2003. Revolutionizing Science and Engineering through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. 84 pp. Available from: http://www.nsf.gov/cise/sci/reports/atkins.pdf

Berman F, Berman J, Pancake C, and Wu L. 2006. A Process –Oriented Approach to Engineering Cyberinfrastructure. Report from the Engineering Advisory Committee Subcommittee on Cyberinfrastructure. 21 pp. Available from: http://director.sdsc.edu/pubs/ENG/report/EAC_CI_Report-FINAL.pdf

Billinge S, Rajan K, and Sinnott S. 2006. From Cyberinfrastructure to Cyberdiscovery in Materials Science. National Science Foundation. Available from: http://www.mcc.uiuc.edu/nsf/ciw_2006/

Committee on Integrated Computational Materials Engineering, National Research Council. 2008. Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security. 152 pp. Available from: http://www.nap.edu/catalog.php?record_id=12199

Committee on Science, Engineering, and Public Policy (COSEPUP). 2007. Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. National Academies Press, Washington D.C. 592pp. Available from: http://www.nap.edu/catalog.php?record_id=11463#toc

Committee on Science, Engineering, and Public Policy (COSEPUP). 2009. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. National Academies Press, Washington D.C. 180 pp. Available from: http://www.nap.edu/catalog.php?record_id=12615

Davis, J. 2006. Cyberinfrastructure in Chemical and Biological Process Systems: Impact and Directions. NSF Cyberinfrastructure Workshop Report. 268 pp. Available from: http://www.oit.ucla.edu/nsfci/NSFCIFullReport.pdf

Deelman E and Gil Y. 2006. Workshop on the Challenges of Scientific Workflows. Report from Workshop on the Challenges of Scientific Workflows: May 1-2, 2006. 22 pp. Available from: https://confluence.pegasus.isi.edu/download/attachments/2031787/NSFWorkflow-Final.pdf?version=1&modificationDate=1254437518000.

DOE Office of Scientific and Technical Information. 2004. The State of Data Management in the DOE Research and Development Complex. Report of the Meeting DOE Data Centers: Preparing for the Future, July 14 – 15, 2004. Available from: http://www.osti.gov/publications/2007/datameetingreport.pdf.

Friedlander A, and Adler P. 2006. To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering. A report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe September 26 – 27, 2006. 159 pp. Available from: http://www.arl.org/bm~doc/digdatarpt.pdf

Interagency Working Group on Digital Data. 2009. Harnessing the Power of Digital Data for Science and Society. Report to the committee on Science of the National Science and Technology council. 24pp. Available from : http://www.nitrd.gov/about/harnessing_power_web.pdf

Lyon, L. 2007. Dealing with Data: Roles, Rights, Responsibilities and Relationships. JISC Consultancy Report. 65pp. Available from: http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf

Lyon L. 2009. Open Science at Web-Scale: Optimizing Participation and Predictive Potential. JISC Consultative Report. 45 pp. Available from: http://www.jisc.ac.uk/publications/reports/2009/opensciencerpt.aspx

National Science Foundation Cyberinfrastructure Council. 2007. Cyberinfrastructure Vision for 21st Century Discovery. 64 pp. Available from: http://www.nsf.gov/pubs/2007/nsf0728/index.jsp

NIOSH [2009]. Approaches to safe nanotechnology: managing the health and safety concerns with engineered nanomaterials. Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2009-125.

NIOSH [2010a]. Progress toward safe nanotechnology in the workplace: A report from the NIOSH Nanotechnology Research Center. Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2010-104.

NIOSH [2010b]. Strategic plan for NIOSH nanotechnology research and guidance: Filling the knowledge gaps. Cincinnati, OH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2010-105.

Novotny MA, Ceperly D, Jayanthi CS, and Martin RM. 2004. Materials Research Cyberscience Enabled by Cyberinfrastructure. Report from The National Science Foundation (NSF) Computational Materials Science Review: June 17—19, 2004. 42pp. Available from: http://www.nsf.gov/mps/dmr/csci.pdf

OECD Environment Directorate. 2008. Working Party on Manufactured Nanomaterials: List of Manufactured Nanomaterials and List of Endpoints for Phase One of the OECD Testing Programme. Series on the Safety of Manufactured Nanomaterials, Number 6.

PCAST. 2010. Report to the President and Congress on the Third Assessment of the National Nanotechnology Initiative. Available from: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nano-report.pdf

PCAST. 2010. Designing a Digital Future: Federally-funded Research and Development in Networking and Information Technology. Available from: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf.

Steering Committee on the Role of Scientific and Technical Data and Information in the Public Domain. 2003. The Role of Scientific and Technical Data and Information in the Public Domain: Proceeding of a Symposium. 238 pp. Available from: http://www.nap.edu/catalog.php?record_id=10785

- Complete list of NSF OCI reports and workshops relating to cyberinfrastructure and its impacts: http://www.nsf.gov/od/oci/reports.jsp
- Complete list of JISC publications, by category: http://www.jisc.ac.uk/en/publications.aspx
- Complete list of OECD nanomaterials publications: http://www.oecd.org/department/0,3355,en_2649_37015404_1_1_1_1_1,00.html
- OECD Database on Research into the Safety of Manufactured Nanomaterials: http://www.oecd.org/document/26/0,3343,en_2649_37015404_42464730_1_1_1_1,00.html
- OECD portal to Nanotechnology Resources by Country http://www.oecd.org/countrylist/0,3349,en_21571361_41212117_42325621_1_1_1_1,00.html