Sharing What We Know Pre-Competitively, Or Is it Time for Nano-Informatics?

John Rumble
Senior Vice President
Information International Associates
Oak Ridge TN

INFORMATION INTERNATIONAL ASSOCIATES (IIA)

www.iiaweb.com

Our Journey Today

- What is Informatics
- Why is it important for nanotechnology and EHS issues in nanotechnology?
- How can informatics help nanotechnology?
- What are the challenges in using informatics?

First, Pre-Competitive Sharing

Pre-Competitive

- Conference presentations
- Published papers
- Grey literature papers
- Standards
- Public databases
- Round robin testing
- Product information sheets

First, Pre-Competitive Sharing

Pre-Competitive

- Conference presentations
- Published papers
- Grey literature papers
- Standards
- Public databases
- Round robin testing
- Product information sheets

Competitive

- Patents
- Copyrights
- Internal reports
- Proprietary information and data
- Actual products
- Manufacturing agreements

What is Informatics

- **Informatics** includes the science of information, the practice of information processing, and the engineering of information systems. (circa 1957-67)
- **Informatics** studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. (1994)

(Wikipedia)

What is Nanotechnology Informatics

- The use of the fruits of the Information Revolution to advance nanoscience and nanotechnology

(Rumble 2006)

Informatics is the study and use of

- Database technology
- Modeling and simulation
- Software
- Networking and connectivity
- Data and resource integration
- Visualization
- Advanced algorithms – computational and numerical
- High-performance computation
- Knowledge discovery

What is Nanotechnology Informatics?

- The ability to do nanoscience and nanotechnology much differently than 60 years ago!

What is Nanotechnology Informatics?

- The ability to do nanoscience and nanotechnology much differently than 60 years ago!

Data

- Generation
- Collection
- Management
- Visualization
- As input to models
- For advanced analysis
- For knowledge discovery

Why is Nanotechnology Informatics Important?

- To reduce the time to commercialize nanotechnology
- To understand and predict nano-behavior
- To design and test nanomaterials and nanodevices systematically
- To optimize nano-performance
- To predict and direct service characteristics of nanomaterials and nanodevices

Why is Nanotechnology Informatics Important?

- To improve nanotechnology, especially on a pre-competitive basis

How Can We Use Informatics to the Advantage of Nanotechnology?

- Database technology
- Modeling and simulation
- Software
- Networking and connectivity
- Data and resource integration
- Visualization
- Advanced algorithms – computational and numerical
- High-performance computation
- Knowledge discovery

How Can We Use Informatics to the Advantage of Nanotechnology?

- Database technology
- Modeling and simulation
- Software
- Networking and connectivity
- **Data and resource integration**
- Visualization
- Advanced algorithms – computational and numerical
- High-performance computation
- **Knowledge discovery**

Data and Resource Integration

- Nano research is a small group effort
- 1000s of groups generate and report results
- The aggregated data and resources are of far greater value than just their sum
- New work is being done and reported as an increasing rate
- Data and resource (models, software, visualization tools) are spread across many disciplines

Data and Resource Integration

- Some disciplines now have highly integrated resources
  - Genomics: GenBank
  - Protein Structure: Protein Data Bank
  - Astronomy: International Virtual Observatory
- Researchers now or soon will have access to all results and many tools
- Google Scholar; SciFinder; blogs, wikis, listservs link researchers together and help find results

Data and Resource Integration

- Non-specialist users still have difficulty
  - finding data and resources
  - understanding quality
  - knowing terminology
- Multi-disciplinary fields such as nanotechnology face formidable obstacles
- Additional multiplicity of EHS aspects of nanotechnology make things even harder
- Is a National Nano Data Center needed?

Knowledge Discovery

- Not just finding new fundamental insights:
  - $E = MC^2$
- Most discoveries and advances are incremental, not *de novo*
- Consider materials design

Materials Design is Design

- **Rules of design apply**
- Data requirements differ by design type
- Computerized data can be reused

Types of Design

- **Redesign** - modify an existing design
- **Routine design** - use a series of rules
- **Mature design** - reuse of the same design every time
- **Selection** - choosing from a catalog
- **Configuration design** - how to assemble
- **Parametric design** - find values from characterizing parameters
- **Original design** - develop something that does not exist

(from Ullman)

Parametric Materials Design

Finding parametric equations for performance from microscopic and atomic models

- Especially needed for complex phenomena
- Quantitative structure activity relationships

- Can equations describe materials behavior? Of course.
- **Can one equation describe many types of materials behavior ???**
- **In the virtual world, can try full range of parameters**

Materials Redesign

Adapting or changing an existing material

- **Most important** type of materials design
- Often just to improve one property
- To be done well need considerable information
- **In the virtual world, can try many changes**

Original Materials Design

Designing something that is totally new

- When do we do original design?
- How often do we start from the beginning?
- Can we define functionality and then start?
- **Origina**l does not equal **originality**
- What we usually mean is **creative design**
- **In the virtual world, we can play**

Back to Knowledge Discovery

- Clearly in the context of nanomaterials, indeed, all of nanotechnology, knowledge discovery from large integrated resources will reduce time and effort for physical experimentation and testing

Back to Knowledge Discovery

- Clearly in the context of nanomaterials, indeed, all of nanotechnology, knowledge discovery from large integrated resources will reduce time and effort for physical experimentation and testing
- For EHS, especially important as human and animal testing may not always be possible
- All combinations far exceeds our capacity to measure

Our Challenges to Using Informatics

- Data collection and preservation
- Data maturity
- Standards
- Data accessibility
- Motivation

Data Collection and Preservation

- Multi-disciplinary nature of nanotechnology makes it difficult to create large collections of nano-related data
- IIa Workshop in January 2007 at Oak Ridge explored some of the issues about a National Nano-Information Center
- No central repositories exist for most nano-related data, especially larger data sets that cannot be published in archival literature

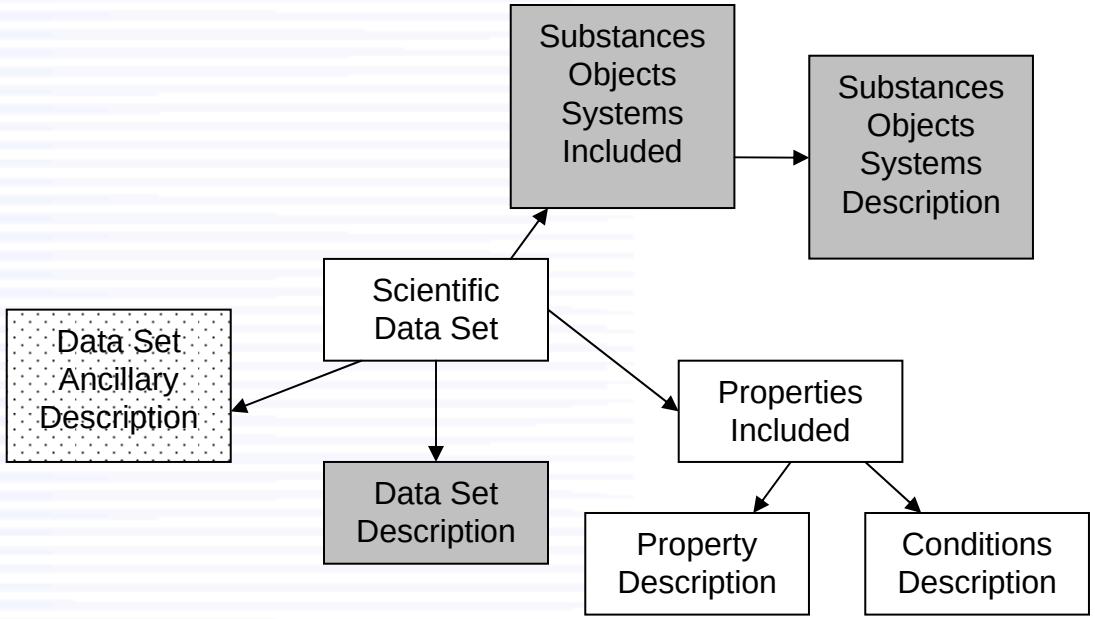
Data Maturity

- Most nanotechnology research is still immature and exploratory
- By immature is meant that many independent variables – in composition, processing, reactivity – are still unknown
- Measurement technology is still changing significantly
- Property data not yet stable

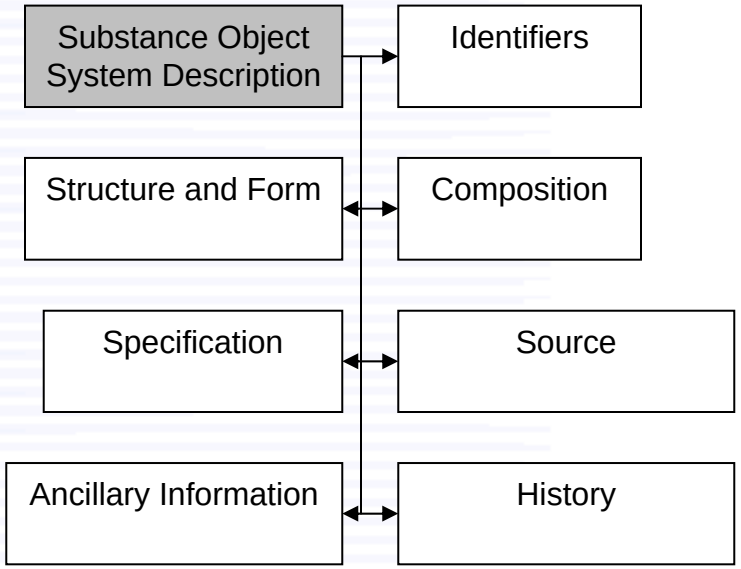
Standards for Nanotechnology Data

- No good nomenclature system
  - Beyond chemical bonding
  - Macroscopic nomenclature is not always appropriate
- Goals on material description
  - Uniqueness – differentiate a material from every other nanomaterial
  - Equivalency – two nanomaterials are the same to degree specified and data sets can be combined

Scientific Data



Components of Materials Description



Standards for Nanotechnology Data

- Test procedures still evolving
- EHS issues are very test dependent
  - Good models/data collections/knowledge discovery help reduce need for physical testing
  - Nanomaterials and nanoparticles interactions with biological materials extremely complex
  - Many Standards Development Organizations are moving ahead

Federated Searches and Data Accessibility

- Searching across data resources very difficult because lack of standards
- Especially true for interactions between nanomaterials and biological materials
- Long term experience with nanomaterials and nanoparticles (e.g., pigments) not easily characterized in nano terminology

Data Accessibility

- Need for nano data collections on a non-proprietary basis
- Pre-competitive data and resource sharing are needed
- Extremely complex science, not easily characterized exactly
- National competitiveness

Nano Informatics

- The tools of the Information revolution have changed how experimental, theoretical, and calculational science is done
- Nano informatics can greatly speed up the development and commercialization of nanotechnology
- Provides alternate methods for addressing EHS issues
- Standards will play a significant role
- National coordination will be needed

National Nano Information Center

- Modeled after National Center for Biotechnology Information
- Repository for pre-competitive information
- Private-public partnership
- Special emphasis on data set completeness, standards and data quality
- Can be a virtual center, with federation of interested parties
- Speed up commercialization and increase national competitiveness