Data Management Plans
A Perspective from the Materials Database Community

Steve Freiman
Freiman Consulting

John Rumble
Information International Associates, Inc.

National Science Foundation

July 28, 2011

Today's Talk

Setting the Stage

- What is materials data?
- Who produces materials data and why?
- Who produces materials databases and why?

What Have We learned About Materials Data and Databases

- Why is there a problem?
- The complexity of materials data
- The role of standards

What lies ahead

What is Materials Data?

- Chemical: What is present in what amounts and how associated
- Structural: How are things arranged; crystals, micro
- Intrinsic: Properties stemming from the structure and composition; not test dependent
- Performance: Test dependent properties to understand how a material performs under specified conditions (stress, cyclical loads, impact, wear, etc.)
- Reactive: Interactions with the environment (corrosion, oxidation, flammability
- Transformation into a product: Ability to be shaped, formed, etc.

Each data category must be considered separately

Who Produces Materials Data and Why?

- Material researchers: For understanding
- Material developers: For a new material
- Materials adaptors: To make an existing material better in some way
- Potential material users: To understand if a material is suitable for some use
- Failure analyst: To understand why a material failed
- Test Developer: To correlate predicted performance better with actual performance
- Regulators: To ensure the correct material is used, or a material is used correctly
- And more

Each use requires a different kind and amount of data

Who Produces Materials Databases and Why

- Government agencies: NIST, DOE, DOD, others
- Professional Societies: ACerS, ASM Intl, JCPDD
- Commercial Publishers: Various plastics handbooks

And Why?

- To support specific types of uses (processing, material selection,analysis)
- To make money
- To provide data sharing for their constituents
- To avoid remeasurements

What Have We Learned?

- That can help us understand what data need to be preserved and how
- That will help NSF develop robust materials data management plans

First a couple of problems

- Stars, crystals, etc. are forever; they are immutable
- Materials are not!!!
  - E.g. Ceramics are complex, rapidly changing, and lack specifications

Materials Evolve

- Collecting all data on things that don't change is relatively easy
  - Just keep doing it and eventually you can pull everything together
- Ceramics change, often faster than one can keep track of the changes
- Today's materials and tests are not the same as yesterday's
- Lack of specifications, e.g. all 96% aluminas are likely to have different properties

The Complexity of Materials

What changes?

- Microstructure
- Composition
- Processing history
- Purity
- Surface treatment
- Heat treatment
- Color; shape; form; etc.
- Workability; formability; joining characteristics
- More

The Complexity of Materials

- When materials change, properties change
- The tests that generate property data also change
- Data compilations are always out-of-date
- Data on the latest material are often proprietary and not available

- Fracture mechanics data on glasses collected at NIST (then NBS) in early 1980’s
- We now know that many of the test procedures used are suspect; data of questionable value

The Role of Formats

- Because of the complexity and varying nature of materials, many types of materials data are difficult to place in standard formats.

Property	**Type**	**Difficulty in Standardizing Data**
Chemical:	What is present in what amounts and how associated	For specific material types (alloys, ceramics, etc) OK; For composites and plastics, can be very difficult
Structural:	How are things arranged	Crystallographic data are well standardized
Intrinsic:	Properties stemming from the structure and composition; not test dependent	Well-defined intrinsic properties lend themselves to standard formats
Performance:	Test dependent properties to understand how a material performs under specified conditions (stress, cyclical loads, impact, wear, etc.)	Can be very difficult depending on complexity of test method, number of variations, standardization of test machines, etc.
Reactive:	Interactions with the environment (corrosion, oxidation, flammability	So far have limited attempts because of complexity of describing environments
Transformation into a product	Ability to be shaped, formed, etc.	For some types (e.g. casting, machinability) OK; others are difficult

What Lies Ahead?

Questions to be considered

- Which types of materials data should be preserved?
- Are there existing materials data repositories?
- What kind of new repositories are needed?
- If not repositories, what?
- What kinds of information for each materials data type needed to make preserved data useful?
  - For the next research step?
  - For transitory use (a decade or less)?
  - For a long time (more than a decade)?
  - Forever?

What Lies Ahead

- One set of repositories clearly won't work
- New repositories require a significant investment. Where will the support come from?
- Not all data are equal. How do you decide what is important enough to preserve?
- What metadata and data fields are needed?
- Metadata standards can help; need three types
    1. The minimum set of metadata necessary to make a data set understandable
    2. How to report other metadata likely to be reported
    3. Format for new types of metadata

Summary

- The complexity of materials and the existence of disparate materials and property types has complicated the development of large scale databases, in contrast to other sciences
- Regardless of the problems, more materials data needs to be preserved than is now being done
- The development of NSF materials data management plans must recognize one size does not fit all types of materials data
- NSF should build on the knowledge gained from more than 30 years of building materials databases and data standards