Accelerating R&D Outcomes Through Data Sharing: Astronomy and Materials Science Examples

Robert Hanisch
Director, Office of Data and Informatics
Material Measurement Laboratory
National Institute of Standards and Technology

Thursday June 13, 2019
Data Sharing in Astronomy

• Long tradition of open data, “open skies”
  – Hubble Space Telescope
  – Sloan Digital Sky Survey
  – National Optical Astronomy Observatory
  – National Radio Astronomy Observatory

• Community-wide common data format: FITS (Flexible Image Transport System)
  – Emerged in late 1970s
  – Became international standard in 1980s
  – Remains in wide use today
Beginning ~2000, international effort started to build distributed but federated data system
  - Data held at major facilities, archive centers, with curation expertise
  - Agreed metadata and protocols for data discovery, access, and interoperability

National programs integrated through International Virtual Observatory Alliance
• Astronomy community developed consensus on discovery, access, and interoperability through the Int’l Virtual Observatory Alliance
  – Bottom-up
  – Minimum required metadata
  – Agreed query and response syntax and semantics

• Data access protocols started simple, grew in capability and complexity as demand increased
  – “cone” search, image access, spectrum access, table access, ...
Virtual Observatory

- More than 10,000 resources (catalogs, image and spectral data collections, etc.)
- Federated search
- Data delivered with standard XML metadata wrapper
- Data providers need not change anything in their data systems, only provide a common interface layer
• Efficiency
• Easy comparison of data from diverse resources
  – Radio, infrared, optical, ultraviolet, x-ray
  – Catalog cross-matching
  – Correlation studies
  – Discovery of statistical outliers
• VO-enabled research highlights at http://ivoa.net
• Astronomy was FAIR before FAIR existed
Hubble Space Telescope Publications

- Totals: 6515 GO, 6527 AR, 2394 Combined
- ~60% of peer-reviewed publication rely on archival data!
Materials Genome Initiative

- Human Welfare
- Clean Energy
- National Security
- Next Generation Workforce

Materials Innovation Infrastructure

- Computational Tools
- Experimental Tools
- Digital Data
• Astronomy data
  – Limited number of data types
  – No commercial interests

• Materials science data
  – Plethora of data types, experimental methods, simulations
  – Highly customized instrumentation
  – Complex sample processing histories
  – Proprietary data formats
  – Major commercial interests
• NIST initiated metadata schema for resource discovery
• Established International Materials Resource Registry Working Group within the Research Data Alliance
• Used framework of RDA to build international consensus
• Implemented in Materials Resource Registry with multiple installations (distributed but federated)
LIMS Help Manage the Data Lifecycle

Laboratory Information Management System

- Acquire
- Process
- Analyze
- Store
- Share
- Reuse
- Dispose
- Plan
- Read + Extract
- Front-End File Management Tools
- Convert + Export
- Archive
- Metadata
- Data

NIH, June 14, 2019
• Capture instrument metadata at the source
  – Metadata extractors
  – Often must reverse engineer proprietary binary formats
• Move experiment metadata into database
  – Enable search across many experiments
  – Do not rely on filenames/file system for metadata storage
• Enable scripted data processing, calibration, feature extraction
• Support data management from acquisition to publication; improve reproducibility
Apple watch
-Announced September 2014

Baseline: 316L Stainless Steel
- Cold-forged to 40% harder
- Special purity mirror finish

High Strength 18K Gold
- 2X harder

Milanese Loop Alloy
- Custom Magnetic Stainless Steel

Anodizable 7000 Aluminum
- 60% stronger Al
- 30% lighter than 316L

c/o P. Voorhees, Northwestern University
Questek’s Ferrium M54 Steel for Hook Shank Application

From design to deployment in 8 years using the MGI approach

Certification of alloy was required


c/o P. Voorhees, Northwestern University
Corning Gorilla Glass 3 for Consumer Electronics

- Gorilla glass used in 4.5 billion devices
- Goal: design a new a stronger thinner Gorilla glass
- No material certification needed, hence even faster to market

From design to deployment in 22 months

Materials Innovation Case Study, [https://www.nist.gov/mgi/mgi-reports](https://www.nist.gov/mgi/mgi-reports)

c/o P. Voorhees, Northwestern University
Take Aways

• Sharing of well-curated data enables new science and more efficient science, increases return on investment, and increases transparency and reliability of research outputs
• Effective sharing requires community engagement in data and metadata standards
• Need to assure that researchers get credit for data sharing