Open Data and Data Citation at NIST

Robert Hanisch
Office of Data and Informatics
Material Measurement Laboratory
National Institute of Standards and Technology
• Strengthen NIST’s commitment to providing public access to scientific research results
• Support governance of and best practices for managing peer-reviewed scholarly publications and digital scientific data across NIST
• Ensure effective access to and reliable preservation of NIST peer-reviewed scholarly publications and digital scientific data for use in research, development, education, and scientific discovery
• Increase use to NIST research results to enhance scientific discovery, education, and research and development across the US
• Enhance innovation and competitiveness by maximizing the potential to create new business opportunities

http://www.nist.gov/open/
Data Categories

- **SRD**: Defined by Standard Reference Data (SRD) Act
- **Reference Data**: Like SRD but not distributed under SRD Act Authority
- **Resource Data**: Data that we’ve used to underpin a position or decision
- **Published Results**: Results that are published in a paper or on a website
- **Publishable Results**: Results that are of sufficient quality that they could be published
- **Derived Data**: Raw data that has been analyzed or processed in some way
- **Working Data**: Raw observational data coming off of an instrument

Hanisch, Data Citation Workshop, July 12, 2016
NIST peer-reviewed papers published on or after 1 October 2015 must be put into PMC within a year of publication. Technical reports go to GPO FDsys.

http://www.ncbi.nlm.nih.gov/pmc/funder/nist/
NIST publications will be discoverable through PMC, FDsys, CHORUS, nist.gov, Google Scholar, and publisher websites.
Materials Science Data Repository

Communities in NIST Repositories

Select a community to browse its collections:
- ASM Structural Materials Data Demonstration Project
- CHiMaD Data Collections
- Computational File Repository
- Experimental Data Repository
- Genome in a Bottle
- Heusler Phases: First Principles Simulations
- ICME Approach to Development of Lightweight 3G AHSS Vehicle Assembly
- ICME of Carbon Fiber Composites for Lightweight Vehicles
- MGI Catalogs
- NanoRelease
- NIST/DOE-EERE Advanced Automotive Cast Magnesium Alloys
- NIST Thermodynamics and Kinetics Test Space
- RDA Demonstration Project: DTR/PID & MGI Infrastructure
- Synchrotron Studies of Slot Die Coated Films
- Thermal Conductivity of CVD Diamond - DARPA Round Robin
- TMS Springer Integrating Materials and Manufacturing Innovation (IMMI)

Recently Added

Al-Cu Symmetric/Asymmetric Tilt Grain Boundary Dataset
Tschopp, Mark A.; Coleman, Shawn P.; McDowell, David L.
Symmetric and asymmetric tilt grain boundaries in Cu and Al were generated using molecular statics energy minimization in LAMMPS with in-plane grain boundary translations and an atom deletion criterion. The following ...

http://materialsdata.nist.gov/

Hanisch, Data Citation Workshop, July 12, 2016
Materials Science Data Repository

Persistent Identifier

Offer licenses with attribution 3.0

Hanisch, Data Citation Workshop, July 12, 2016
• Program dedicated to improving public access to NIST research data outputs
• Management of Institutional Data Assets (MIDAS) tool
  – Data management plans
  – Resulting data described in Enterprise Data Inventory (EDI)
  • Project Open Data schema https://project-open-data.cio.gov/v1.1/schema/ with extensions
  • All public-facing NIST data
  • All published data objects assigned DOIs
  • JSON records uploaded to Dept. of Commerce, data.gov
• Data repository and discovery portal
  – Integrated search across all NIST public data assets
  – Likely to use Fedora-based repository
Cite NIST Datasets

Description

Journals often have their own citing format for publications and other items included in article reference lists. NIST’s data citation recommendations are consistent in using the same citation elements (author, title, publisher, year of publication, etc.) as used by journals. In the absence of other (e.g., journal) defined requirements, these recommendations provide a consistent manner for appropriate attribution to scientific datasets that are made available under NIST’s public access policy.

The following minimum requirements and format should be used to cite NIST datasets (a URL is used if a PID is not available):

**Recommended Format:**

Author/editor (Publication Year), Title, Publisher, Persistent Identifier (PID) or URL (Access date)

**Example:**


It may also be desirable to include information from two optional properties, Version and Type of Medium, as appropriate. If so, the recommended format with NIST example is as follows:

**Recommended Format:**

Author/editor (Publication Year), Title, Version, Publisher, [Type of Medium] Persistent Identifier (PID) or URL (Access date)

**Example:**

• Katherine Sharpless, NIST Open Data Officer
• Del Brockett, NIST CIO
• Anita Vanek, Jim Fowler, NIST Office of Information Systems Management
• Gretchen Greene, Ray Plante, Chandler Becker, Casey Hume, NIST Office of Data and Informatics
• Kim Tryka, Andrea Medina-Smith, Regina Avila, NIST Information Services Office (NIST Research Library)
• Jim Warren, Carrie Campbell, NIST Materials Genome Initiative
• And many others...