OSTI Update to ASCAC –

Moving Toward a More "Unified" Scientific Record: Software, Data, Publications

U.S. DEPARTMENT OF ENERGY | Office of Science | osti.gov

Brian Hitson, Director

December 21, 2017

DOE Invests $12B per Year in R&D

U.S. DEPARTMENT OF ENERGY
R&D Funding

NATIONAL LABS

Ames
Argonne
Brookhaven
Fermi
Idaho
Los Alamos
Lawrence Berkeley
Lawrence Livermore
NETL
NREL
Oak Ridge
Pacific Northwest
Princeton
SLAC
Sandia
Savannah River
Thomas Jefferson

GRANTEES

SCIENTIFIC & TECHNICAL INFORMATION (STI/R&D Results)

Text
- Journal articles/accepted manuscripts
- Technical reports
- Conference papers
- Patents

Data
- Large and small datasets
- Images
- Visualizations

Software/Code

≥ 50,000 STI "products" annually

OSTI's mission: make R&D results accessible and useful in the modern science landscape.

P.L. 109-58 (Energy Policy Act of 2005): "The Secretary, ***through the Office of Scientific and Technical Information***, shall maintain within the Department publicly available collections of scientific and technical information resulting from research, development, demonstration, and commercial applications activities supported by the Department."

Review of the ASCAC-STI Subcommittee Charge

- Are OSTI products and services best in class?
- What is the national and international standing of OSTI?
- In what areas must OSTI be a clear leader?
- Recommendations for OSTI's future direction.

ASCAC established STI Subcommittee in early 2015, chaired by Dr. Tony Hey.

ASCAC-STI Subcommittee performed on-site review and issued report (May-Sept. 2015).

Summary of Subcommittee Answers and Recommendations

- OSTI products have a number of “best in class” capabilities
  - SciTech Connect’s semantic search
  - Data ID Service (providing DOIs for datasets)
  - ScienceCinema’s speech indexing for multimedia
  - Federated search engines Science.gov and WorldWideScience.org
  - NOT best in class: Energy Science and Technology Software Center
- OSTI’s “standing” nationally and internationally
  - “OSTI is in a leadership position among Federal agencies”
  - “OSTI services employ a range of innovative technologies not uniformly available from their peer international scientific information organizations”
- OSTI must be a clear leader in providing public access to DOE-funded scholarly publications

Summary of Subcommittee Answers and Recommendations (cont'd)

- While there is significant uptake of OSTI services by the public and commercial services, OSTI needs to have a better understanding of researcher needs and should "initiate a vigorous outreach program with the DOE Lab researchers."
- OSTI needs to "re-invent the ESTSC software service."
- OSTI should continue to work "toward a unified user environment with a limited number of clearly delineated, non-redundant tools."
- DOE and OSTI need to effectively implement public access, address publication content gaps, and incentivize labs and grantees to support DOE's Public Access Plan.
- The Office of Science should consider "defining a useful role for OSTI . . . in managing DOE data."

OSTI Actions and Progress

1. Outreach to DOE research community
   - Four lab workshops in 2016 (ORNL, BNL, SLAC, LBNL)
   - Data and software “roundtables” 2016-2017 (NREL and INL)
   - 9 software re-invention project requirements teams, comprising 51 lab and DOE stakeholders
   - Data ID Service workshops at SLAC (2016) and OSTI (2017)
   - Workshop scheduled at LLNL in March 2018

Feedback incorporated to develop product roadmaps and to improve Data ID Service, DOE Data Explorer, “unified” product features, and new software dissemination model.

OSTI Actions and Progress (cont'd)

2. Reinvent the Energy Science and Technology Software Center (ESTSC) service – DOE's software submission and dissemination tool
   - × Not comprehensive
   - × Not modern

- Defined requirements for a reinventing ESTSC, incorporating feedback from DOE researchers, developers, policy, legal, and technical communities – in all, 9 requirements teams
- Renamed DOE CODE – (launched November 2017) open source code and requirements posted to GitHub.

DOE CODE

DOE CODE repository on GitHub
- github.com/doecode/

DOE CODE landing page
- www.osti.gov/doecode/

DOE CODE

CREATE

SUBMIT

DISCOVER

DOE CODE FEATURES

DOE CODE resources provide you information on best practices and policies for software

DOE CODE is an open source product on GitHub that other institutions can download and deploy for their own purposes

EASY

DOE CODE is easy to use and requires minimal metadata, often autopopulated

DOE CODE seamlessly interfaces with common development platforms (e.g., open web API)

doi®

DOE CODE issues digital object identifiers (DOIs) for code so software is more easily cited and discoverable. Allows for stronger connections between code, publications, and data through citation in reference sections

DOE Software Catalog

- Launched with ~700 projects
- Another 800 candidates on GitHub
- Partnering with DOE CIO to feed Code.gov

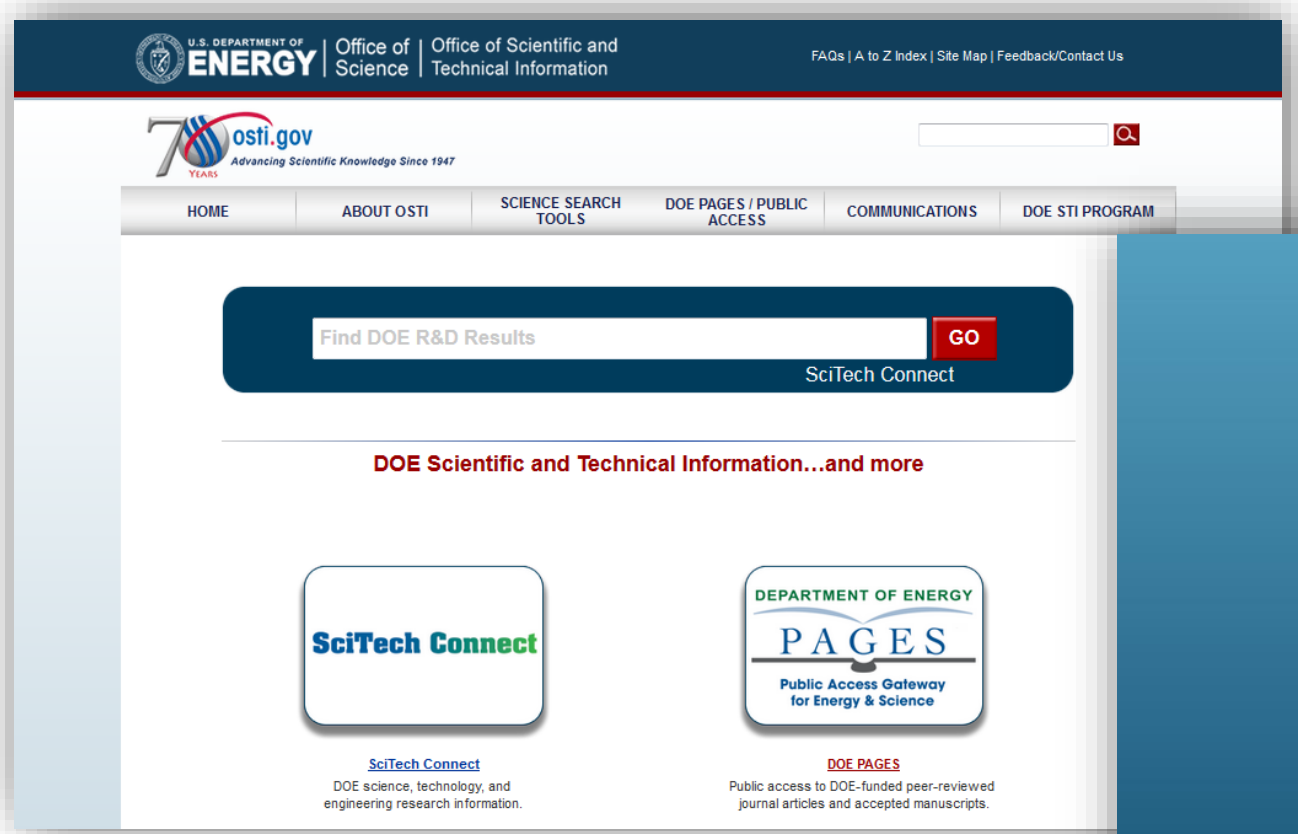
DOE CODE offers public and private repository services in GitHub and GitLab

OSTI Actions and Progress (cont'd)

3. Toward a unified user environment
   - Product streamlining and consolidation
     - Consolidated/eliminated 10 standalone web products since 2014
     - January 2018 consolidation of OSTI.gov website and umbrella product SciTech Connect to make OSTI.gov synonymous with search of DOE R&D results.

OSTI.GOV

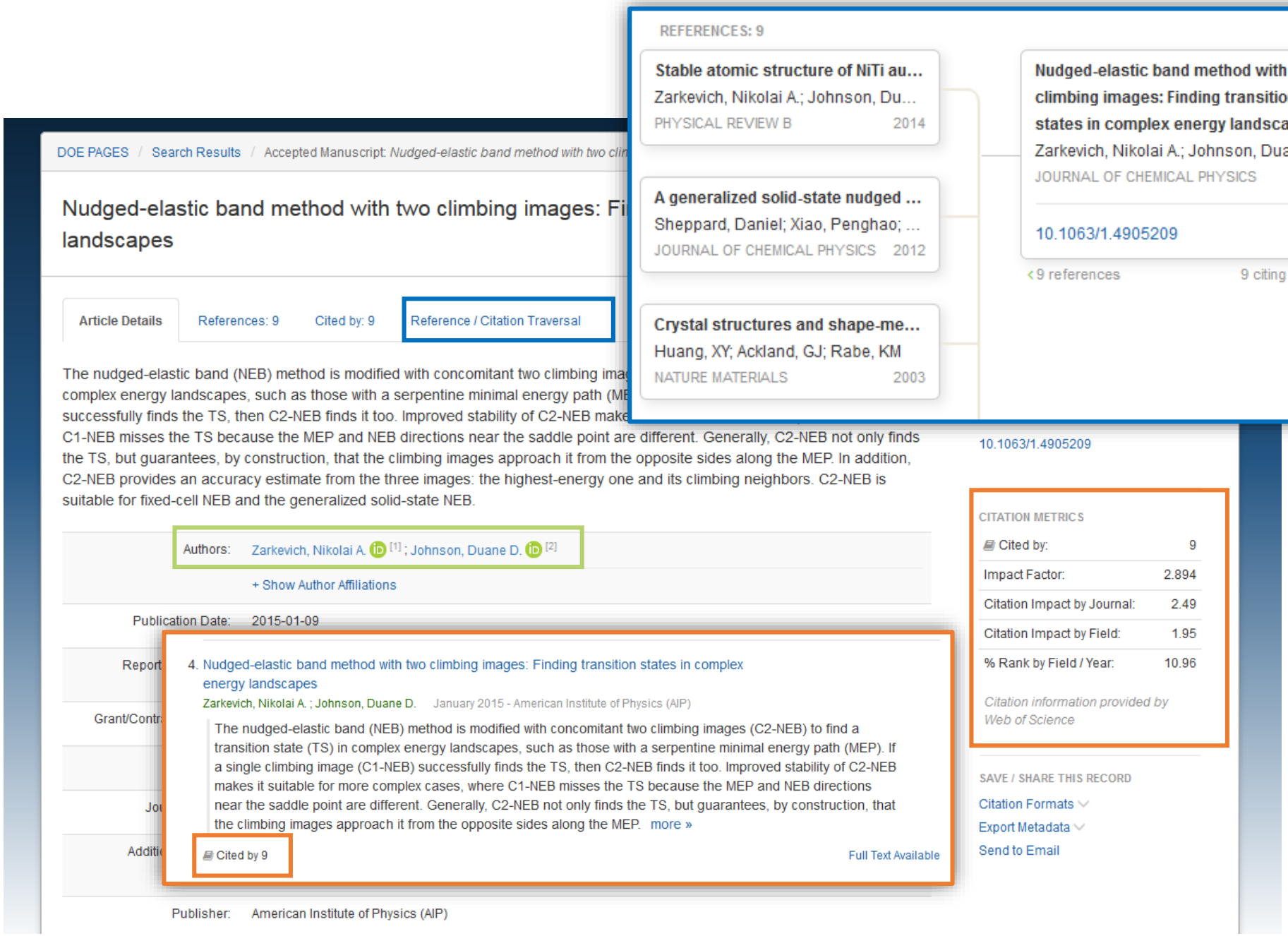
OLD → NEW


U.S. DEPARTMENT OF ENERGY | Office of Science | Office of Scientific and Technical Information
FAQs | A to Z Index | Site Map | Feedback/Contact Us
osti.gov
Advancing Scientific Knowledge Since 1947
HOME
ABOUT OSTI
SCIENCE SEARCH TOOLS
DOE PAGES / PUBLIC ACCESS
COMMUNICATIONS
DOE STI PROGRAM
Find DOE R&D Results
GO
SciTech Connect
DOE Scientific and Technical Information...and more
SciTech Connect
DEPARTMENT OF ENERGY
PAGES
Public Access Gateway for Energy & Science
SciTech Connect
DOE science, technology, and engineering research information.
DOE PAGES
Public access to DOE-funded peer-reviewed journal articles and accepted manuscripts.

OSTI.GOV
U.S. Department of Energy
Office of Scientific and Technical Information
Search 2.9+ million Department of Energy research results
Submit Research Results
Search Tools
Public Access Policy
Data Services & Dev Tools
About
FAQs
News
U.S. DEPARTMENT OF ENERGY | Office of Science | Office of Scientific and Technical Information
Website Policies / Important Links
Contact Us

OSTI Actions and Progress (cont'd)

3. Toward a unified user environment
   - Product streamlining and consolidation
     - Consolidated/eliminated 10 standalone web products since 2014
     - January 2018 consolidation of OSTI.gov website and umbrella product SciTech Connect to make OSTI.gov synonymous with search of DOE R&D results
   - Improved product features responsive to researcher feedback

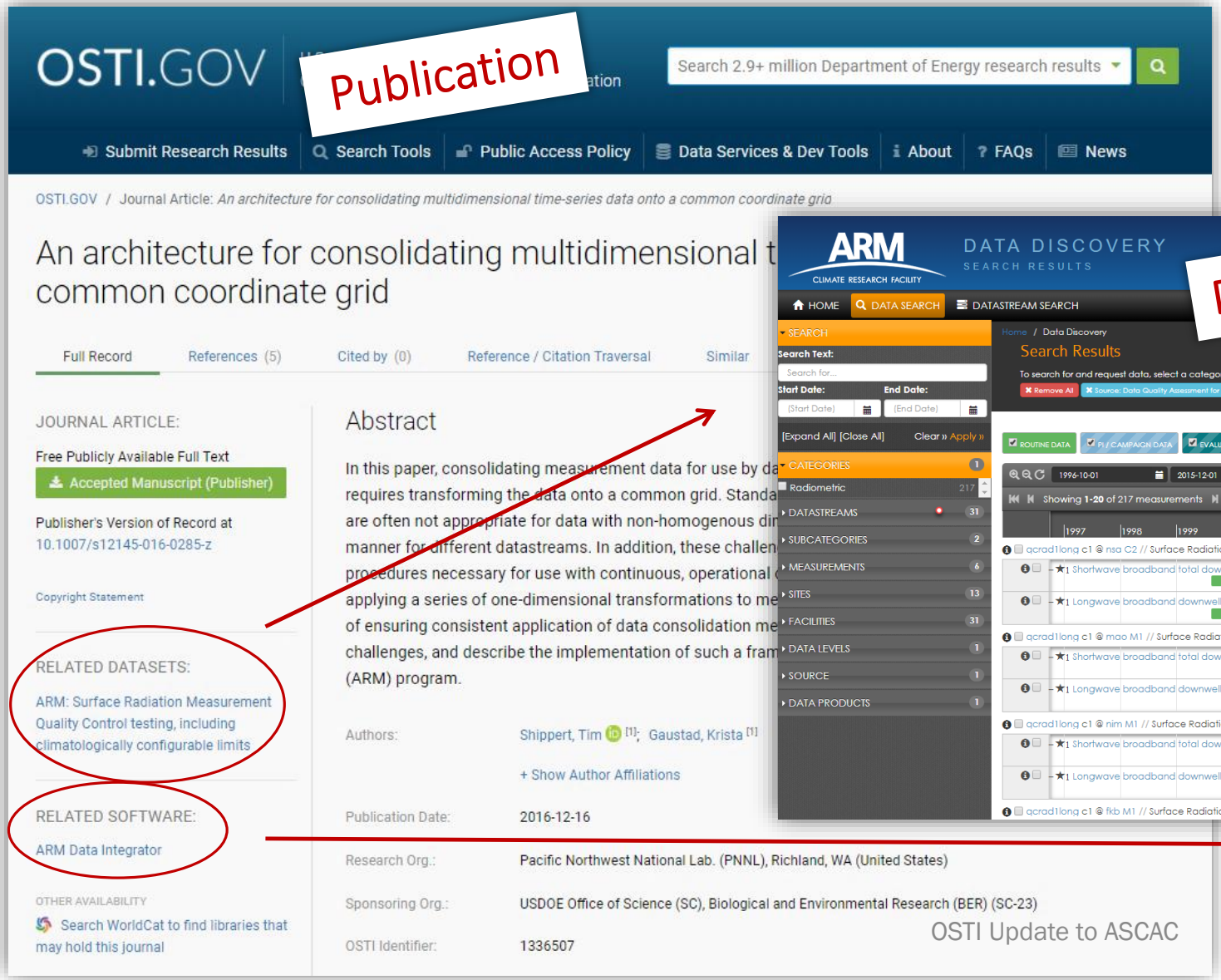


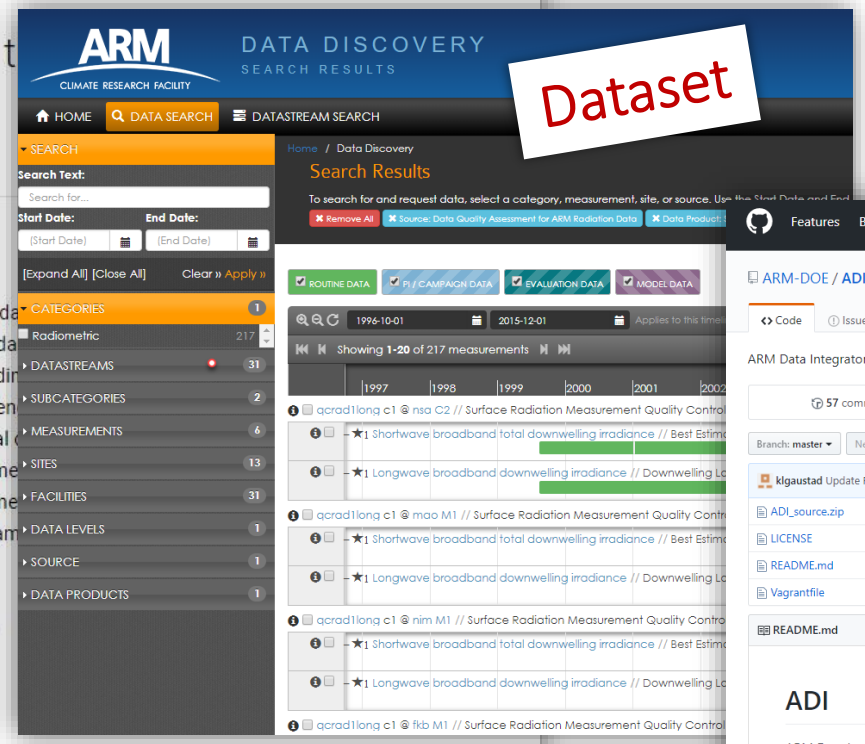
New Product Features

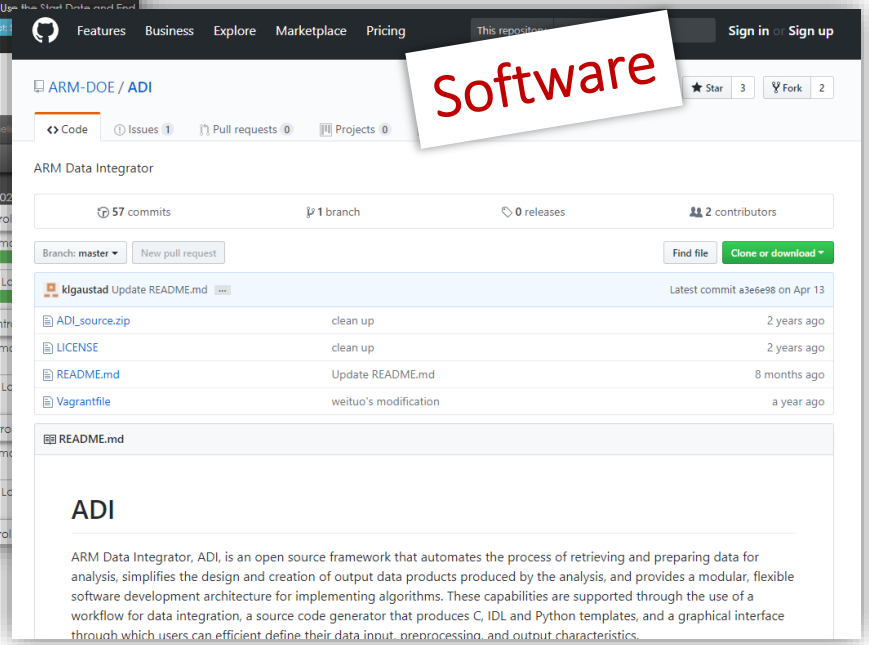
- Reference/Citation Traversal
- Author Profiles/ORCiDs
- Citation Metrics

Modern Science Demands Reproducibility

End state goal: interlink all related research outcomes (e.g., from publication to related data to related software).





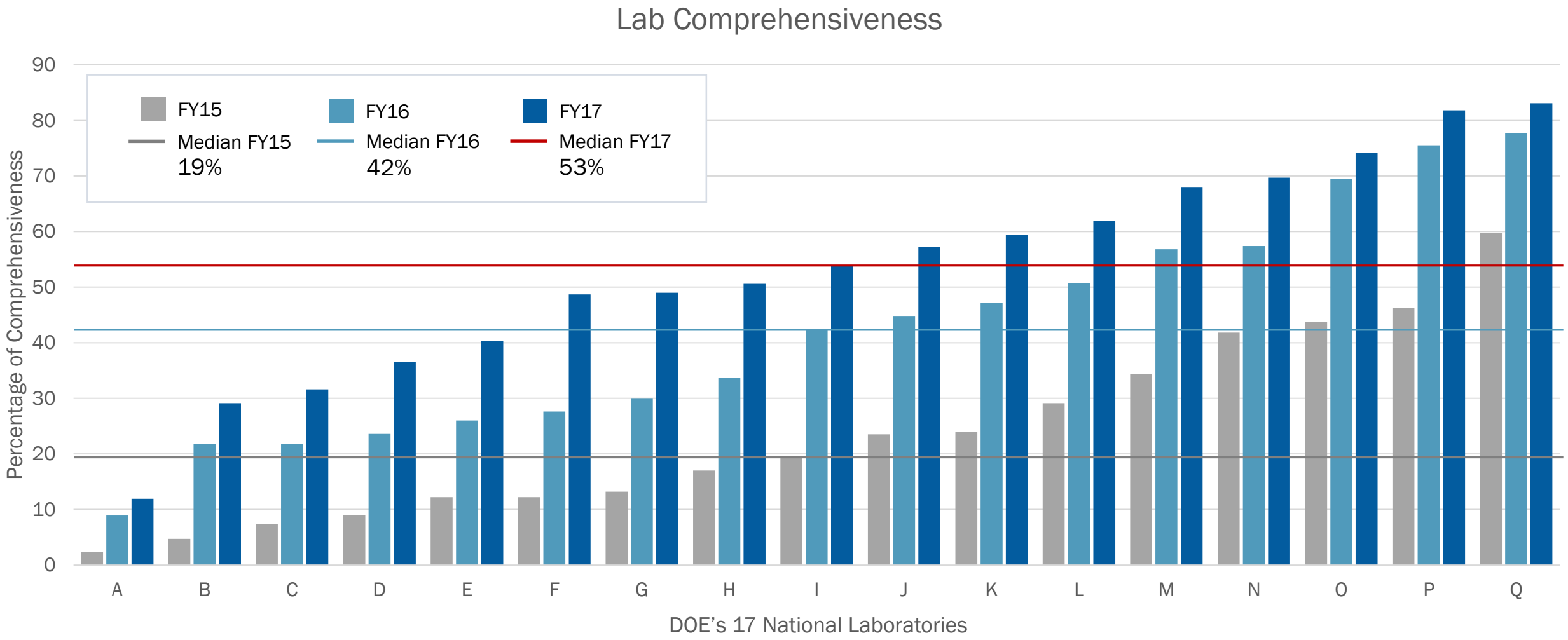


OSTI Actions and Progress (cont'd)

4. Effectively implement public access
   - Address publication content gaps. OSTI uses Web of Science to establish "denominator" for articles authored by lab researchers, and labs collect accepted manuscripts.
   - Incentivize labs and grantees to support DOE public access efforts. SC and all DOE program offices established public access as a measurable in labs' annual performance plans.

https://youtu.be/mZky964Lef8

Public Access Progress



OSTI Actions and Progress

5. OSTI's role in DOE's data landscape
   - Primary contribution is the DOE Data ID Service:
     - Enabled by OSTI's membership in DataCite
     - Issued >70,000 DOIs to 21 DOE data clients
     - Makes datasets discoverable through OSTI products and indexing by Google and other common search engines
     - OSTI also provides this service to 7 other federal agency data clients on cost-reimbursable basis
   - Implementing supplemental material submission with publications in 2018
   - Exploring image, graph, table extraction from publications. Investigating both automated and human curation techniques; possibly a candidate for machine learning
   - Workshop feedback identified needs for data repository services; DMP guidance and best practices; and hierarchical relationships among datasets

Conclusions

As a result of ASCAC-STI recommendations and support from SC and individual subcommittee members, OSTI:

- Initiated ongoing, active outreach efforts to DOE lab research community to inform product improvements;
- Modernized scientific software dissemination model with launch of DOE CODE;
- Further unified and streamlined OSTI product environment, moving toward increased interlinking of research objects;
- Increased DOE's comprehensiveness in providing public access to scholarly publications through DOE PAGES;
- Broadened the DOE Data ID Service to improve discoverability of DOE datasets.

OSTI appreciates SC and ASCAC support and is committed to continued progress.

Thank you,

Brian A. Hitson, hitsonb@osti.gov