Open Science from the U.S. Government

And U.S. Department of Energy Implementation

U.S. DEPARTMENT OF ENERGY

2019 DOE Data ID Service Workshop • September 17-18

DOE DATA ID Service

History of Public Access and Open Science in U.S. Government

2009 NIH Public Access (FY2009 Omnibus Appropriations Act); PubMed Central

2013 OSTP Memo to Federal Agencies; implement Public Access Plan to increase access to scholarly publications and digital data

2013 OMB M-13-13 Open Data Policy; required federal data to be available in machine-readable formats

2014 FY2014 Omnibus Appropriations Act codified OSTP memo into law for HHS, Labor, CDC, Education

2016 OMB M-16-21 Federal Source Code Policy; required custom-developed federal source code to be made available for reuse (Code.gov)

2016-Present National Science and Technology Council -> Committee on Science -> Subcommittee on Open Science

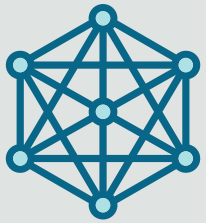
Results of U.S. Federal Research and Development

$150B in federal R&D

~200,000
journal articles
per year

~7,000
patents
per year

Exabytes
of data



>6,500
open source
software projects
(Code.gov)

Why Be Open?

1 Accelerate scientific progress

2 Promote collaboration

3 Enable reproducibility

the "paper" plus "code, data, etc. that can be used to reproduce the results." (Wikipedia)

3 Pillars of Open Science

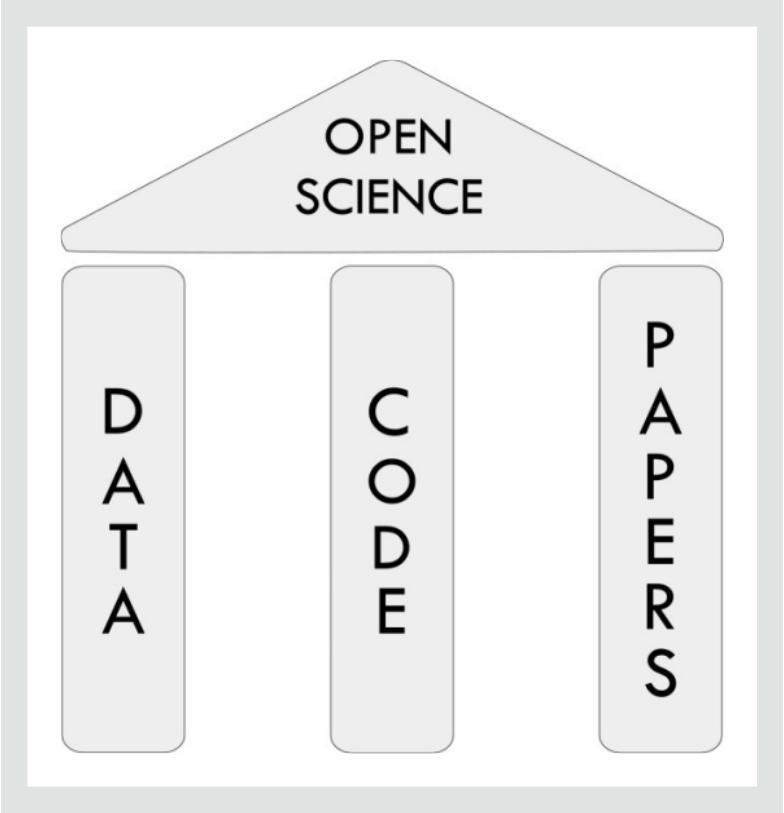
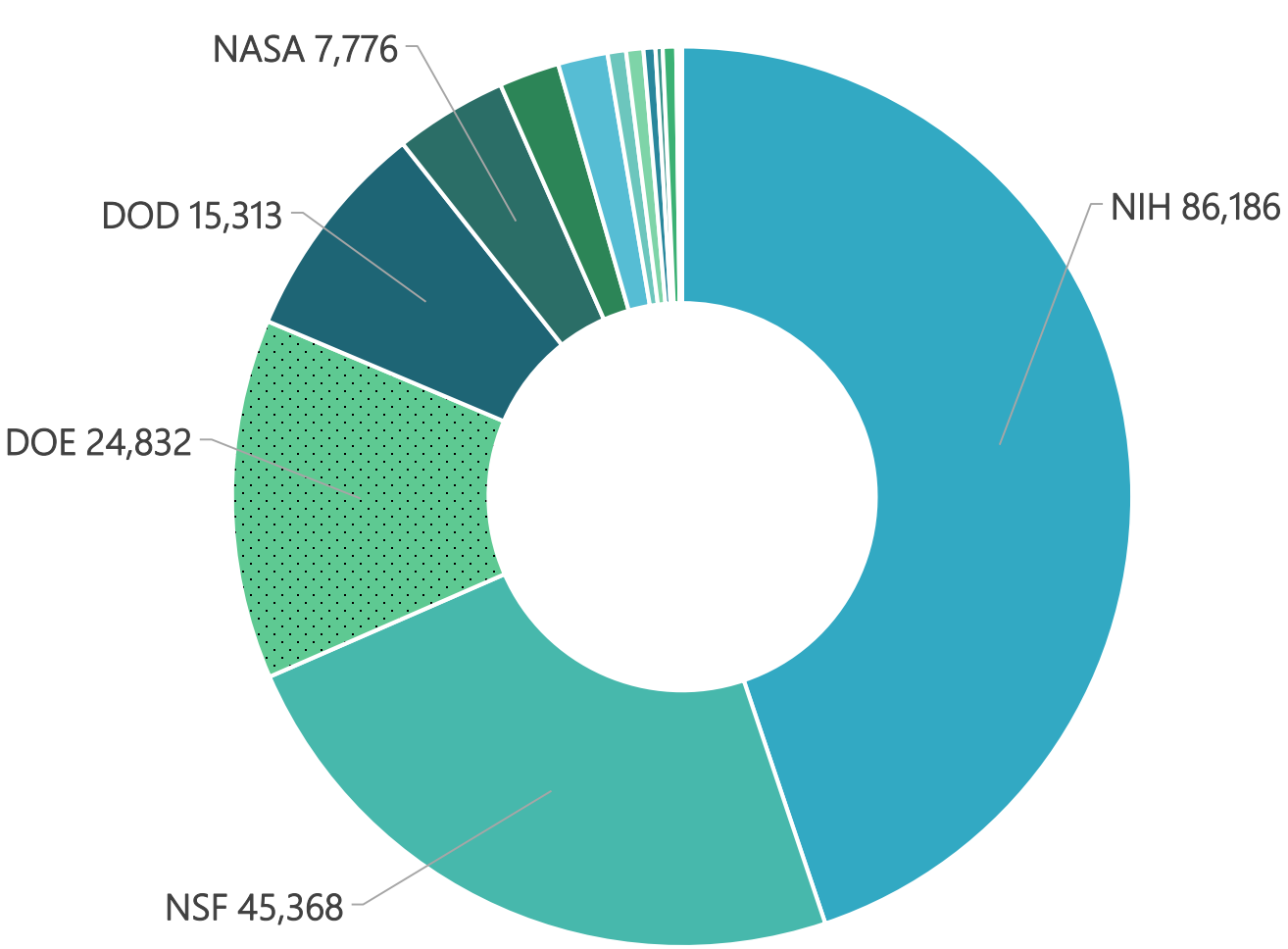


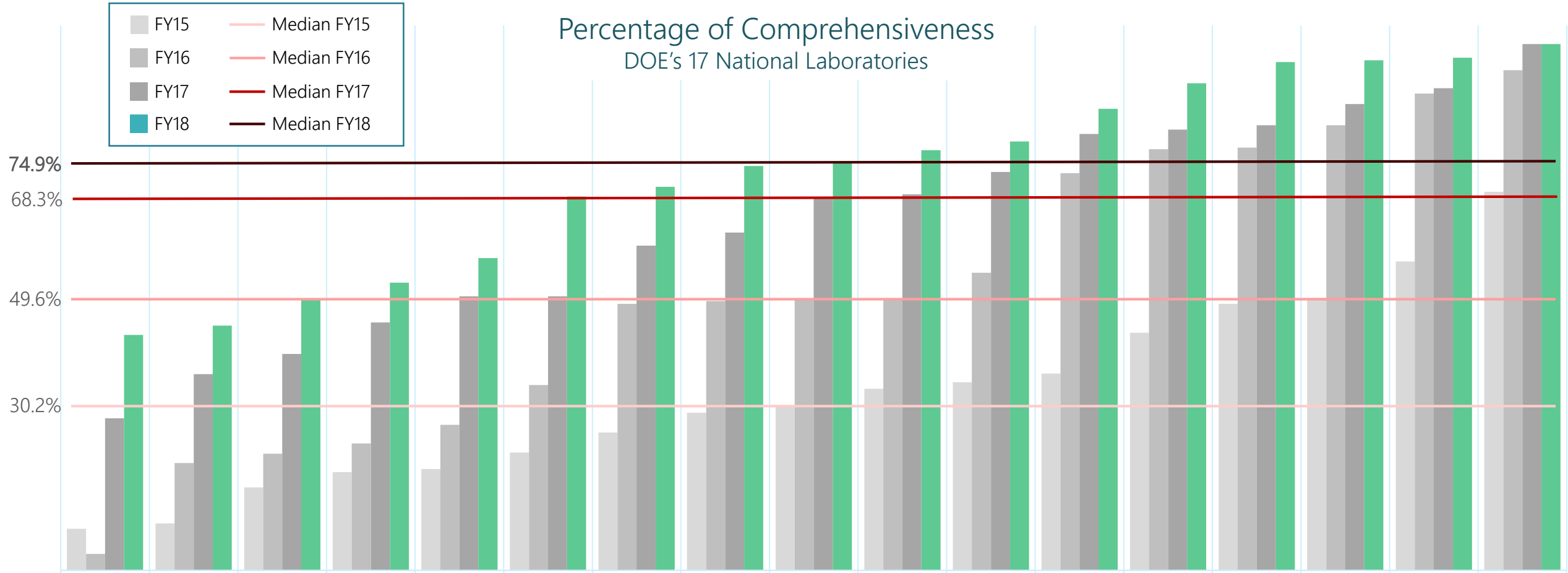
Photo source: PLOS Biology, "A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research," Krzysztof J. Gorgolewski, Russell A. Poldrack, 2016

Article/Manuscript Output by Agency CY2018



- National Institutes of Health
- National Science Foundation
- Department of Energy
- Department of Defense
- National Aeronautics and Space Administration
- Department of Commerce
- Department of Agriculture
- Department of Veterans Affairs
- Centers for Disease Control
- Environmental Protection Agency
- U.S. Geological Survey
- U.S. Fish & Wildlife Services
- Department of Homeland Security
- Department of Education
- Department of Transportation

DOE's Public Access Progress Since 2015



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
FY15	7.6	8.6	15.2	18.0	18.6	21.6	25.3	28.9	30.2	33.3	34.5	36.1	43.6	48.9	50.0	56.7	69.5
FY16	3.0	19.7	21.4	23.3	26.7	34.0	48.9	49.4	49.6	49.8	54.6	72.9	77.3	77.6	81.7	87.5	91.8
FY17	27.9	36.0	39.7	45.5	50.3	50.3	59.6	62.0	68.3	69.0	73.1	80.1	80.9	81.7	85.6	88.5	96.6
FY18	43.2	44.9	49.8	52.8	57.3	68.6	70.4	74.2	74.9	77.1	78.7	84.7	89.4	93.3	93.6	94.1	96.6

But . . . There's Much More than Publications . . .

U.S. DEPARTMENT OF ENERGY

R&D Funding

$12B per year

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

The Office of Scientific and Technical Information (OSTI) mission: make R&D results accessible and useful in the modern science landscape.

DOE DATA ID Service

How Do We Add Data and Software to Open Science Landscape?

- Follow FAIR principles (Findable, Accessible, Interoperable, Reusable)
- FAIR requires persistent identifiers (e.g., DOIs) and good metadata
- DOE integrates third-party persistent identifier services from
  - DataCite to obtain DOIs for data and software
  - Crossref to obtain DOIs for gray literature and to auto-populate metadata for journal articles
  - ORCID for author services and searching
- Research community has to practice and expect data and software citation (e.g., cite/link to underlying data and software in papers)

How DOE Makes R&D Results Accessible and Useful

Primary Search Tool

OSTI.GOV

www.osti.gov
all DOE R&D

Specialized Search Tools for Specific Resources

DEPARTMENT OF ENERGY
DOE PAGES
Public Access Gateway for Energy & Science

www.osti.gov/pages
journal articles/
accepted manuscripts

DOE Data Explorer

www.osti.gov/dataexplorer
datasets

DOE CODE

www.osti.gov/doecode
software

DOE Patents

www.osti.gov/doepatents
patents

DOE ScienceCinema

www.osti.gov/sciencecinema
video

DOE DATA ID Service

DOE Data Explorer

U.S. Department of Energy
Office of Scientific and Technical Information

Data Discovery Tool
www.osti.gov/dataexplorer

biomass buffers

Explore DOE Data

Explore Projects

A data Project is a specific research group, data center, user facility, or other DOE-funded endeavor that is creating research data.

Explore Collections

A data Collection is a package of related datasets with a DOI for the entire Collection.

Explore Datasets

A Dataset is a single instance of data whose boundaries have been defined by the data creator, with a DOI associated.

DOE Data Explorer

U.S. Department of Energy
Office of Scientific and Technical Information

Search Scientific Data

Data ID Services | About | FAQs | News

DOE Data Explorer / Search for biomass buffers SrRh2 *(filtered)* / Page 1 of 61

609 Datasets

612 Total Results

Sorted by Relevance | Save Results | Prev | ...

SEARCH FOR:

KEYWORDS

biomass buffers SrRh2 [×]

[× clear all] [modify this search]

REFINE BY:

CREATOR / AUTHOR

- Persson, Kristin (241)
- Johnson, Megan (25)
- Uria-Martinez, Rocio (25)
- O'Connor, Patrick (19)
- Chen, Jiquan (11)

more

RESEARCH ORGANIZATION

- LBNL Materials Project (241)
- Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States) (241)
- DOE Joint Genome Institute (89)
- AmeriFlux (55)
- Environmental System Science Data Infrastructure for a Virtual Ecosystem (45)

more

PROJECTS

- The Materials Project (241)
- AmeriFlux Management Project (55)
- Environmental System Science Data

Projects (2) | Collections (1) | Datasets (609)

1. Materials Data on SrRh2 (SG:227) by Materials Project

Persson, Kristin

Computed materials data using density functional theory calculations. These calculations determine t electronic structure of bulk materials by solving approximations to the Schrodinger equation. For mor information, see https://materialsproject.org/docs/calculations

DOI: 10.17188/1189497 | View

2. Tropical Africa: Land Use, Biomass, and Carbon Estimates for 1980 (updated for the year 2000) (NDP-055)

Brown, Sandra ; Gaston, Greg ; Beaty, T. W. ; ...

This document describes the contents of a digital database containing maximum potential abovegrou biomass, land use, and estimated biomass and carbon data for 1980. The biomass data and carbon estimates are associated with woody vegetation in Tropical Africa. These data were collected to redu uncertainty associated with estimating historical releases of carbon from land use change. Tropical A defined here as encompassing 22.7 x 10E6 km ^{2} of the earth's land surface and is comprised of count are located in tropical Africa (Angola, Botswana, Burundi, Cameroon, Cape Verde, Central African Repu Chad, Congo, Benin, Equatorial Guinea, Ethiopia, more »

DOI: 10.3334/CDIAC/LUE.NDP055

3. Leaf Area, Vegetation Biomass and Nutrient Content, Barrow, Ala - 2013

Sloan, Victoria ; McGuire, David ; Euskirchen, Eugenie ; ...

This dataset consists of measurements of vegetation harvested from Areas A to Next-Generation Ecosystem Experiments (NGEE) Arctic site near Barrow, Alaska.

DOE Data Explorer

U.S. Department of Energy
Office of Scientific and Technical Information

Search Scientific Data

Data ID Services | About | FAQs | News

DOE Data Explorer / Search for biomass buffers SrRh2 *(filtered)* / Dataset: *Materials Data on SrRh2 (SG:227) by Materials Project*

Materials Data on SrRh2 (SG:227) by Materials Project

Dataset | Associated Project | Associated Collections (0) | Other Related Research

DATASET:

View Dataset

DOI: 10.17188/1189497

SAVE / SHARE:

Export Metadata

Save to My Library

Abstract

Computed materials data using density functional theory calculations. These calculations determine the electronic structure of bulk materials by solving approximations to the Schrodinger equation. For more information, see https://materialsproject.org/docs/calculations

Creator(s)/Author(s):	Persson, Kristin
Publication Date:	2015-02-09
Other Number(s):	mp-1328
DOE Contract Number:	AC02-05CH11231; EDCBEE
e:	Dataset
g.:	LBNL Materials Project; Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States)
Sponsoring Org.:	USDOE Office of Science (SC), Basic Energy Sciences (BES) (SC-22)
	6 MATERIALS SCIENCE
	rystal structure; Rh2 Sr1; Rh-Sr; ICSD-105937; ICSD-650413; electronic bandstructure
OSTI Identifier:	1189497
DOI:	10.17188/1189497

DOI: 10.17188/1189497

DOE obtains DOI from DataCite and assigns to this dataset

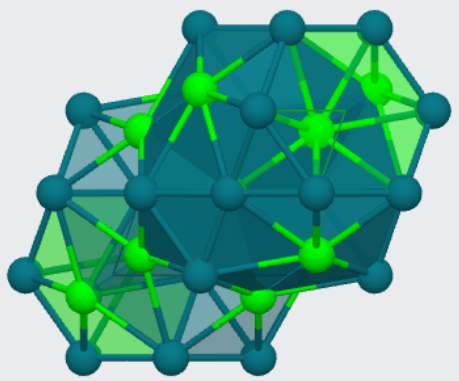
MATERIAL
$SrRh_2$

ID:
mp-1328

DOI:
10.17188/1189497

Show Help Guides

Electronic Structure | X-Ray Diffraction | X-Ray Absorption | Substrates | Elasticity | Similar Structures | Calculation Summary | Provenance/Citation



Rh Sr

☑ Atoms ☑ Unit Cell ☑ Bonds ☑ Polyhedra

CIF

Zoom in/out	Shift + Drag cursor
Rotate along the center axis	Option + Drag cursor

Edit Crystal | Generate Phase Diagram

Tags: Rhodium strontium (2/1)

File Formats | Download

Material Details

Final Magnetic Moment
0.001 μ_B

Magnetic Ordering
NM

Formation Energy / Atom
-0.254 eV

Energy Above Hull / Atom
0.000 eV

Density
8.28 g/cm^3

Decomposes To
Stable

Band Gap
0.000 eV

Space Group

Hermann Mauguin
Fd$\bar{3}$m [227]

Hall
F 4d 2 3 $\bar{1}$d

Point Group
m$\bar{3}$m

Crystal System

Lattice Parameters

a	5.502 Å	α	60.000°
b	5.502 Å	β	60.000°
c	5.502 Å	γ	60.000°

Volume 117.750 Å^3

Final Structure

Fractional Coordinates

CIF

Sr

a	b	c
0.125	0.125	0.125
0.875	0.875	0.875

Rh

a	b	c
0	0.5	0.5
0.5	0	0.5
0.5	0.5	0
0.5	0.5	0.5

Software Discovery Tool
www.osti.gov/doecode

DOE CODE | U.S. Department of Energy Office of Scientific and Technical Information

Search DOE CODE for submitted software entries

Submit Software/Code | Repository Services | Software Policy | Resources | About | FAQs | News

DOE CODE / Search for neutrino

3 Search Results

SEARCH FOR:
KEYWORDS:
neutrino [×]

[× clear all] [modify this search]

REFINE BY:

PROJECT TYPE
- Open Source, Publicly Available Repository
- Open Source, No Publicly Available Repository
- Closed Source, Site Hosted
- Closed Source, OSTI Hosted

LICENSES
- Other (Commercial or Open-Source)
- Apache License 2.0
- GNU General Public License v3.0
- MIT License
- BSD 2-clause "Simplified" License
- BSD 3-clause "New" or "Revised" License
- Eclipse Public License 1.0
- GNU Affero General Public License v3.0
- GNU General Public License v2.0
- GNU General Public License v2.1
- GNU Lesser General Public License v2.1
- GNU Lesser General Public License v3.0
- Mozilla Public License 2.0
- The Unlicense

SOFTWARE TYPE
- Business

1. CosmicEmu, Version 2.0

Lawrence, Earl Release Date: 2017-04-20

The CosmicEmu implements a Gaussian process emulation
matter + neutrino power spectra over a range of physical pa
evolution of the Universe. The procedure can produce thous
the complex n-body simulations used to produce the trainin
training simulations, their design, and the general scheme is
et al. (2014), Heitmann et al. (2016), and an upcoming pape

2. Time Resolved Phonon Spectroscopy, Ve

Goett, Johnny ; Zhu, Brian Release Date: 2017-01-12

TRPS code was developed for the project "Time Resolved Pl
piece of software were specially created to model phonon g
interact with ionizing radiation, particularly applicable to the
dark matter and neutrino research. These routines were crea
Geant4 framework for the modeling of radiation transport in
them for eventual integration into that code base.

3. brulilo, Version 0.x

Malone, Chris Release Date: 2015-04-15

brulilo is a Python package for building and evol
astrophysical explosion calculations. It is very m
functionality is still under development. The mai

DOE CODE | U.S. Department of Energy Office of Scientific and Technical Information

Search DOE CODE for submitted software entries

Submit Software/Code | Repository Services | Software Policy | Resources | About | FAQs | News

DOE CODE / Search Results / CosmicEmu, Version 2.0

CosmicEmu, Version 2.0

Full Project

RESOURCE

Publicly Accessible Repository

https://github.com/lanl/CosmicEmu

DOI: 10.11578/dc.20171025.1933

SAVE / SHARE

Export Metadata

Abstract

The CosmicEmu implements a Gaussian process emulation scheme to predict the dark matter and dark matter + neutrino power spectra over a range of physical parameters that describe the composition and evolution of the Universe. The procedure can produce thousands of spectra per second, much faster than the complex n-body simulations used to produce the training set, which can take weeks per simulation. The training simulations, their design, and the general scheme is described in Lawrence et al. (2010), Heitmann et al. (2014), Heitmann et al. (2016), and an upcoming paper.

Developers:	Lawrence, Earl [1]
	Show Developer Affiliations
Release Date:	2017-04-20
Project Type:	Open Source, Publicly Available Repository
	Scientific
Licenses:	Other (Commercial or Open-Source)
	USDOE
	Primary Award/Contract Number: AC52-06NA25396
	5212
Site Accesion Number:	7455

DOI: 10.11578/dc.20171025.1933

DOE obtains DOI from DataCite and assigns to this software

Why GitHub? Enterprise Explore Marketplace Pricing Search Sign in Sign up

lanl / CosmicEmu

Watch 2 Star 6 Fork 4

Code Issues 1 Pull requests 1 Projects 0 Security Insights

Branch: master CosmicEmu / P_cb / emu.c

Find file Copy path

earlcl Update emu.c 1fa125a on Apr 18, 2017

1 contributor

389 lines (323 sloc) 10.9 KB

Raw Blame History

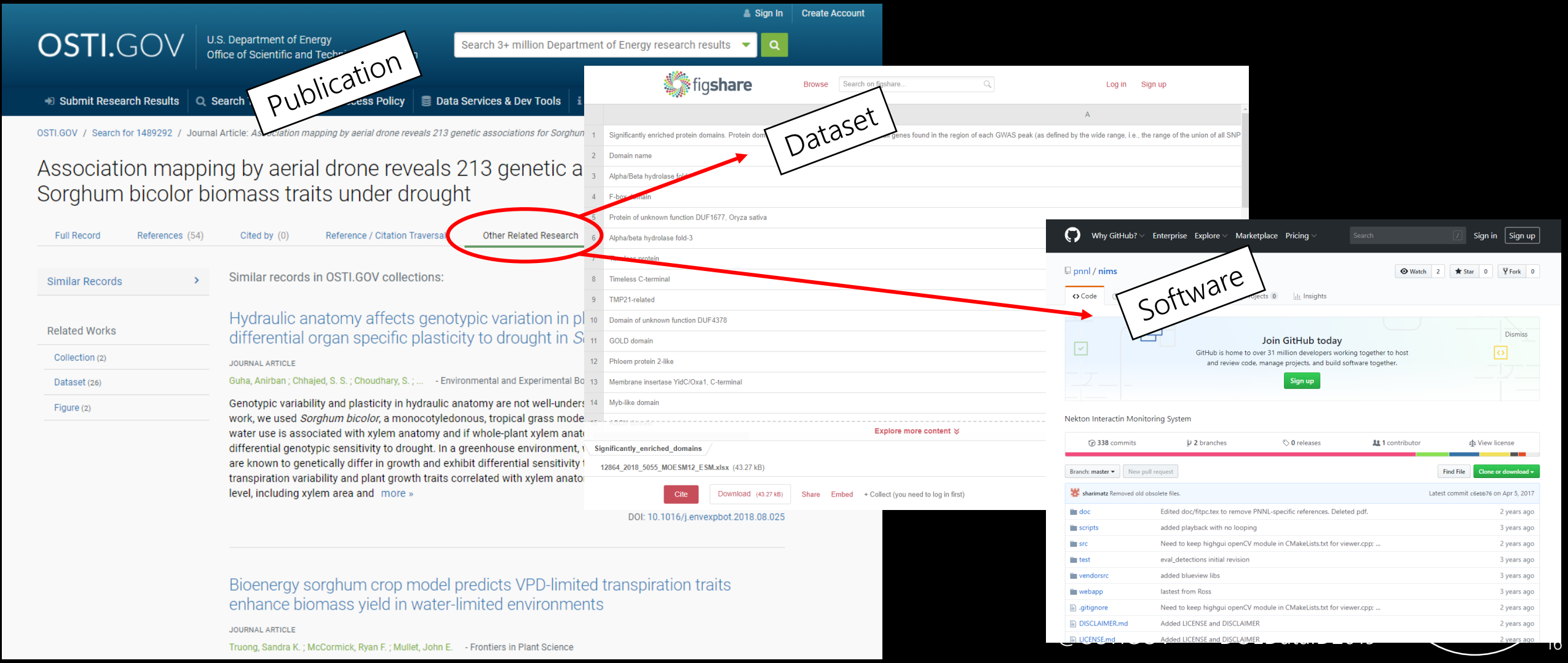
```
//
//  emu.c
//
//
//  Created by Earl Lawrence on 11/10/16.
//
//

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <string.h>
#include <gsl/gsl_matrix.h>
#include <gsl/gsl_linalg.h>
#include <gsl/gsl_spline.h>
#include <gsl/gsl_errno.h>

#include "params.h"
```

Interlinking through DOIs

= open science in practice



Enabling Broader Discovery

DOE "pushes" its R&D results through multiple channels; e.g.,

> Google Dataset Search[Beta] (enabled through DOIs)

> Data.gov; Code.gov (catalogs of U.S. agency collections)

> Federated Search Products

Science.gov – simultaneous search of all U.S. agencies' scientific information

WorldWideScience.org – multilingual search of ~75 national and international databases

Thank you

U.S. DEPARTMENT OF ENERGY

OSTI.GOV
Office of Scientific and Technical Information

Carly Robinson
Assistant Director Information Products and Services
U.S. Department of Energy
Office of Scientific and Technical Information

Carly.Robinson@science.doe.gov https://www.osti.gov

DOE DATA ID Service