

SO YOU WANT TO SHARE YOUR RESEARCH DATA?

The Penn State Data Commons

# WHAT IS THE PENN STATE DATA COMMONS?



The Data Commons is an institutional repository for Penn State research data.

Over 300 PSU researchers share data through the Data Commons.

It also hosts core data to support research, teaching, and outreach.

# BACKGROUND



2005 Concept first presented as an environmental data library and repository data created by PSU faculty, researchers, and collaborators. Received a \$5K grant from PSU to explore the concept and acquire data.



2023 The Data Commons hosts and provides access to approximately 800tb of data. Data storage is provided by the Institute for Computational and Data Sciences. Institute for Energy and the Environment supports DC staff and EMS has staff member who supports data transfer to DC.

# SERVICES

---

Data discovery & public access

---

Data storage

---

Data archiving

---

Support for development of agency required data management plans

---

Metadata development and data management seminars

---

Creation of Digital Object Identifiers (DOIs)

---

Visualization of compatible data

---

Core data

---

# APPROACH



Mediated approach—no self service.



Why a mediated approach?



We work directly with researchers throughout the data management life cycle.



Data is often too large to upload via self service mechanisms.



Develop more complete metadata.



Need to compress and organize data for download or create streaming services.

# WHAT DOES THE DATA COMMONS PROVIDE ACCESS TO?



Research data



Core Data: imagery,  
databases,  
spreadsheets,  
geospatial data,  
satellite imagery,  
vector data, grids)



Metadata



Documentation of  
protocols (included in  
metadata)



Models



Apps & Tools



Reports



Publications (via links,  
DOIs)

# WHO SHARES DATA VIA DATA COMMONS?

Meteorology

Civil  
Engineering

Chemistry

Aerospace  
Engineering

Forest  
Resources

Acoustics

Ecosystem  
Science

Animal  
Science

Mechanical  
and Nuclear  
Engineering

Mining  
Engineering

Entomology

Landscape  
Architecture

Geography

Meteorology

Physics

# WHO SHARES DATA VIA DATA COMMONS?

Huck

EESI

IEE

Sustainability  
Institute

SAFES Institute

Noid Lab

Center for  
Environmental  
Informatics

Ice and Climate  
Research  
Center

# PROJECTS

---

NewBio

---

Shalenetwork

---

Ensemble Mars Atmosphere Reanalysis System

---

Flood Mitigation in Pennsylvania's Rural Communities

---

Piloting a Data and Metadata Hierarchy for Use around United States Cities and Airports

---

Earth system and climate model output

---

High-Resolution Rapid Refresh (HRRR) Model

---

VORTEX

---

Atmospheric Marcellus shale measurements

---

National Carbon Program's Mid Continental Intensive (MCI)

---

Indianapolis Flux (INFLUX) experiment

---

WRF-GHG model output from North American domain

---

Terrestrial Lidar Data

## OTHER INITIATIVES

Multiple internal seed grants also rely on being able to access and store data via the Data Commons including the SSRI Reducing Social Disparities, IEE, and other seed grants.

In addition, the Data Commons has enabled organizations at Penn State such as the Arboretum, Office of the Physical Plant, and the Sustainability Institute to collect and display data for use in applications like the PSU Arboretum Plant Finder, PSU Arboretum Commemorative Finder, the Tree Database, and Lion Ambassador's Tailgate Recycling initiative.

# BUT I HAVE BIG DATA, CAN YOU HANDLE IT?

Big data is data that is one or all of the following: high volume, high velocity (speed of data in and out), thousands or millions of files.

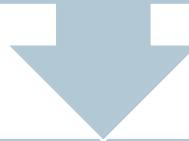
Data Commons hosts small data sets and big data.

We have expertise in working with large, complex data.

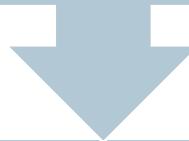


# IMPACT

The presence of the Data Commons at Penn State positions the university as a leader in providing public access to research data.

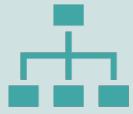


It demonstrates Penn State's ability to support big data projects and programs and enables Penn State to be even more competitive for large multi-institutional data driven projects.



Data Commons aligns to the FAIR principles, and is an approved repository for multiple journals, RE3Data.org, and the Research Data Alliance, and data.gov

# GETTING STARTED: NEW PROPOSALS



In your data management plan, include the Data Commons. We have information you can use to populate your DMP.



Be sure to consider funding for managing and storing your data. Data Commons is hosted by the Institute for Computational and Data Sciences (ICDS). The funding should support ICDS.



Feel free to reach out to us during your project. We can provide advice on Best Management Practices for data management for future publication.

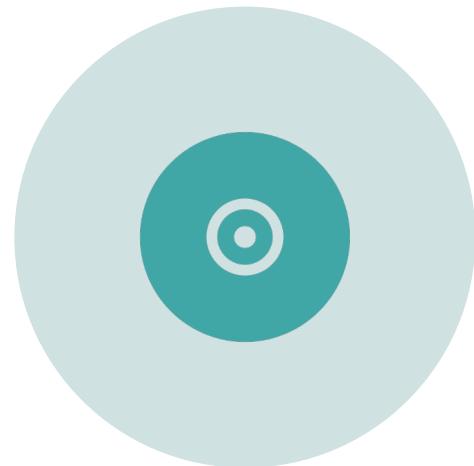
# READY TO SHARE?

Contact us at [datcommons@psu.edu](mailto:datcommons@psu.edu)

We will work with you to get your data ready to share publicly.

We will create metadata, a DOI, and links to related material.

# THANK YOU!



SITE:  
[DATACOMMONS.PSU.EDU](http://DATACOMMONS.PSU.EDU)



EMAIL:  
[DATACOMMONS@PSU.EDU](mailto:DATACOMMONS@PSU.EDU)