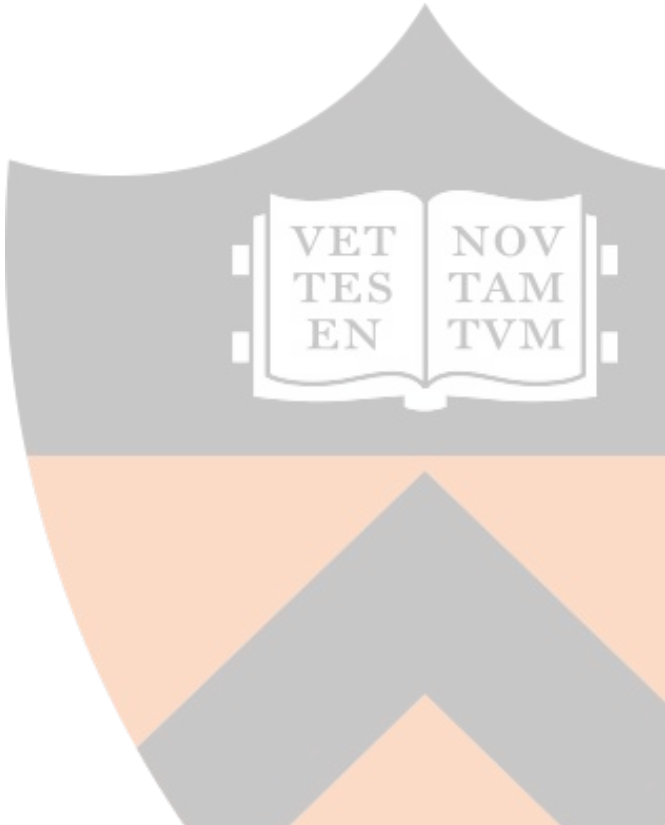
Agility, Growth, and Cooperative Service Design

Two teams building a custom data repository during times of change

Kate Lynch (they/them)
Lead Library Software Engineer
Princeton University
katelynch@princeton.edu

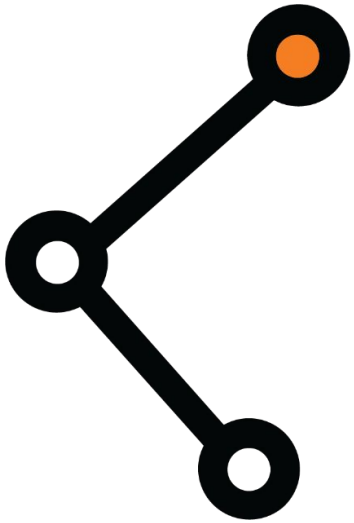
Meghan Testerman (she/her)
Open Research and Scholarship Librarian
Head, Princeton Research Data Service
Princeton University
mtesterman@princeton.edu



Introductions

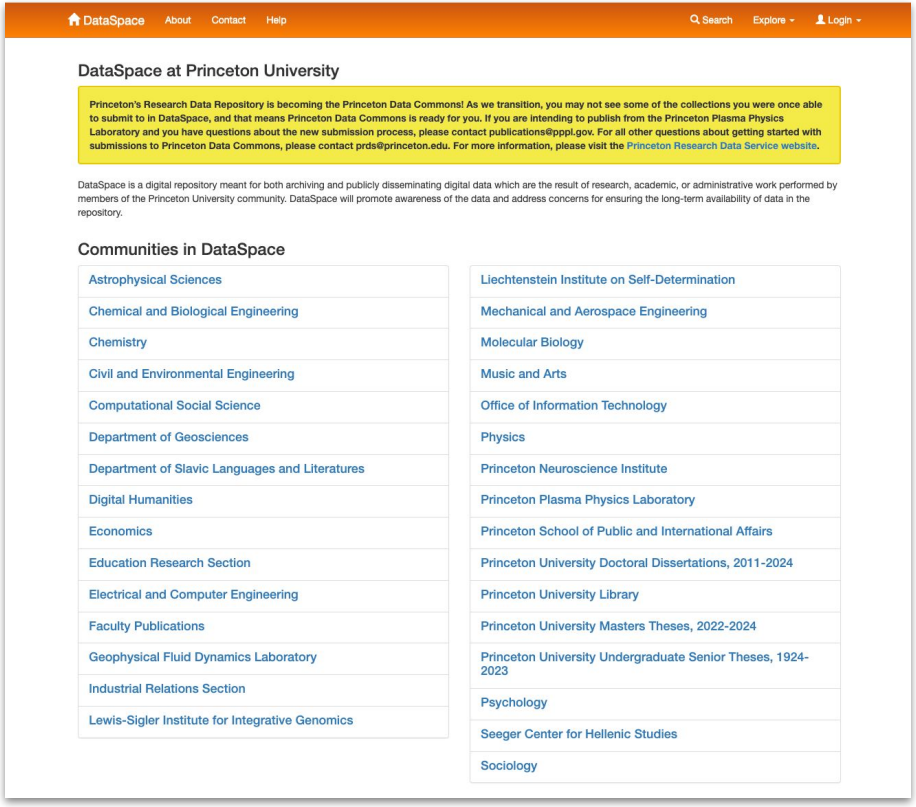
Princeton University Library

Princeton Research Data Service

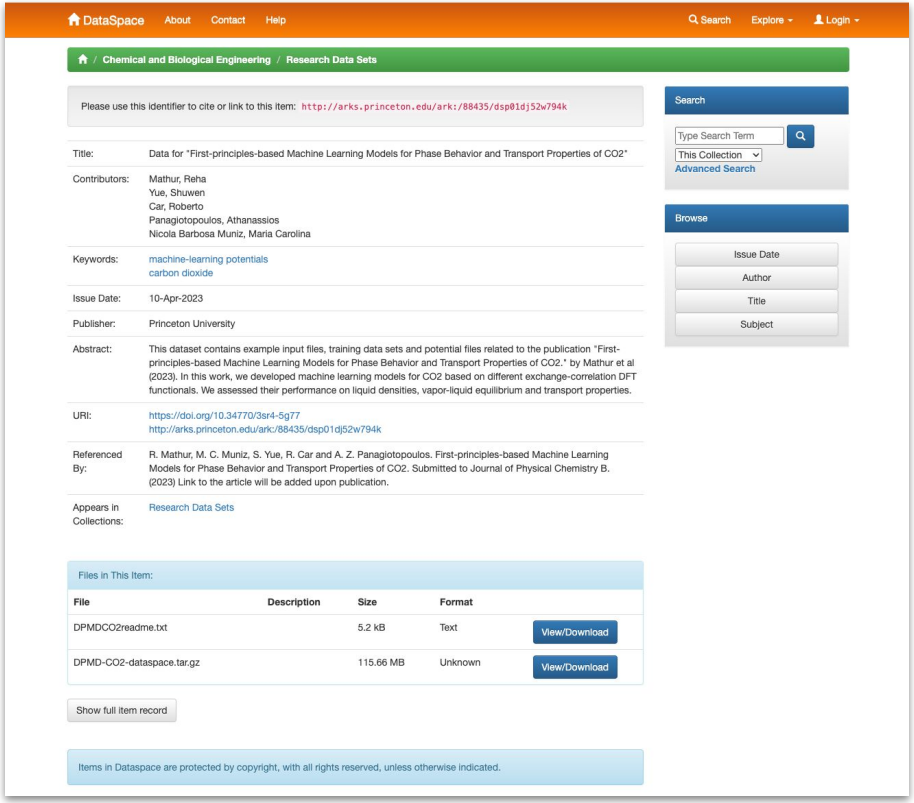


Princeton Data Commons

How we got here - DataSpace

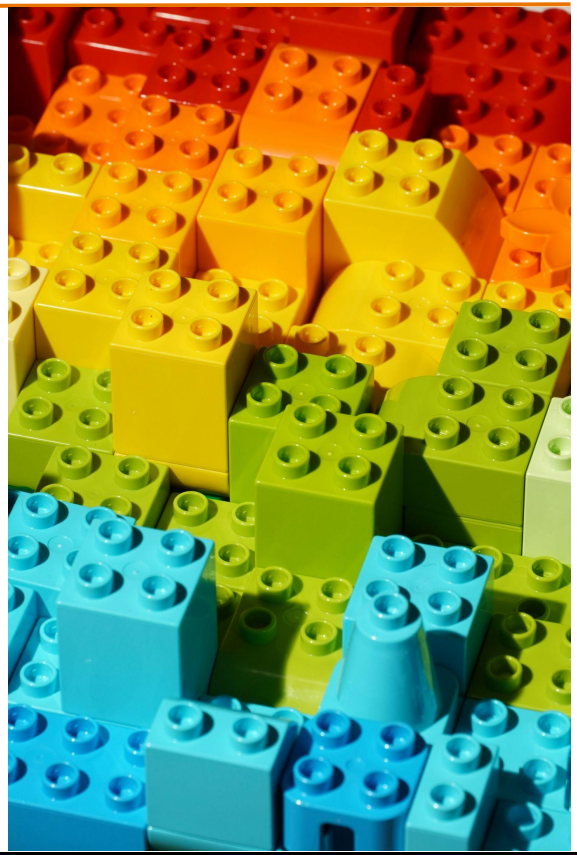


How we got here - DataSpace



The solution

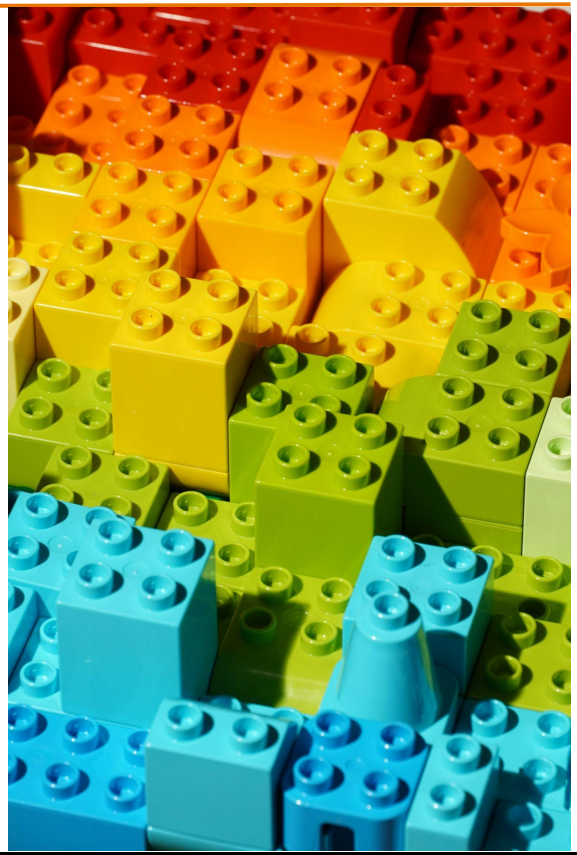
- **Princeton Data Commons (PDC)**
  - Blacklight and Ruby on Rails
  - Code repositories
    - https://github.com/pulibrary/pdc_discovery
    - https://github.com/pulibrary/pdc_describe



The solution

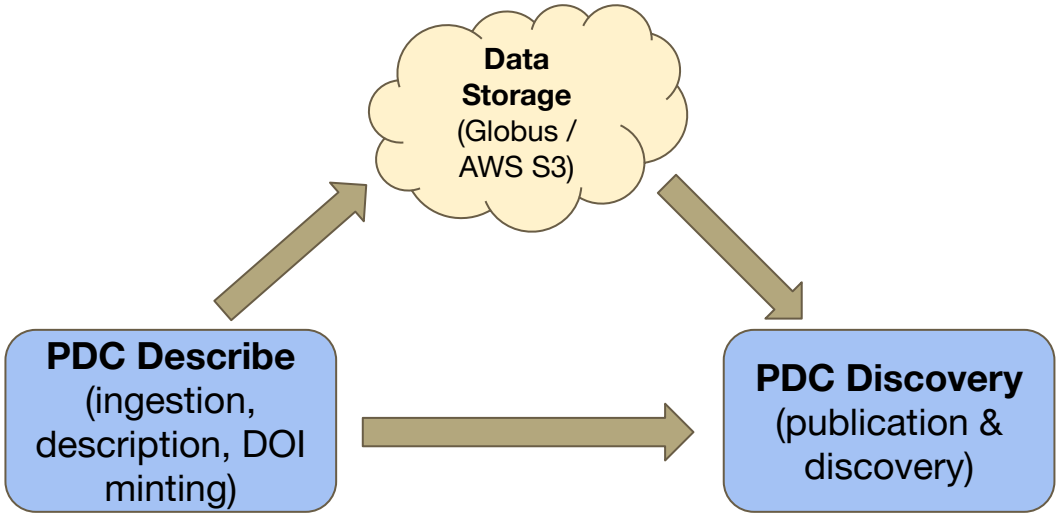
- **Princeton Data Commons (PDC)**
  - Blacklight and Ruby on Rails
  - Code repositories
    - https://github.com/pulibrary/pdc_discovery
    - https://github.com/pulibrary/pdc_describe

- Sustainable software, ability to add and enhance features, integrate with other services developed by RDSS for Research Data



Princeton Data Commons

DataSpace
(ingestion, ARK minting, publication & discovery, data storage)



Princeton Data Commons - Features

Data curation: Fully mediated by a team of 3 curators

Access Options: immediately available or embargo of up to 1 year

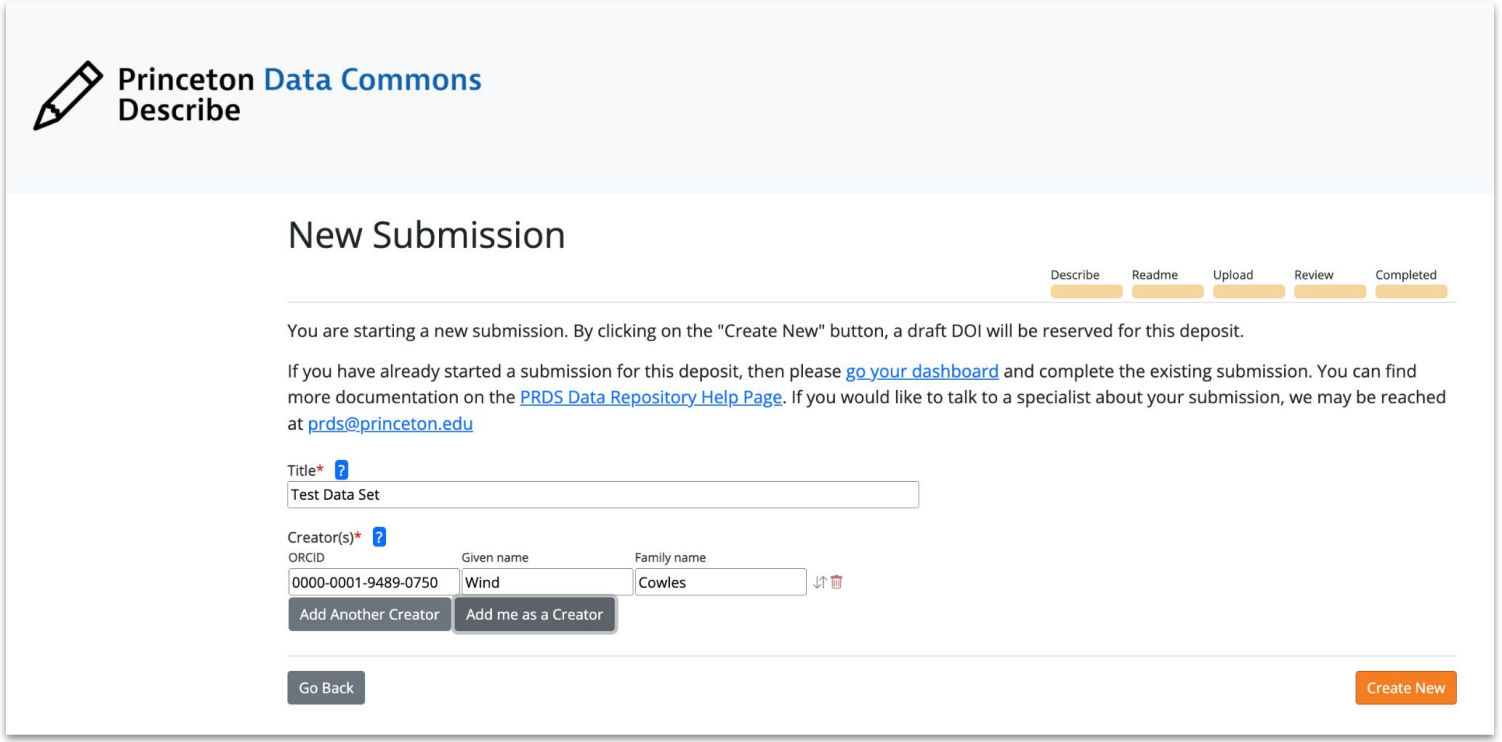
Discovery and Access: Immediate access to by download or request to all data not under embargo

Persistent Identifiers: DOI / persistent link

Large data: Globus for uploading, curation, and downloading of data

Meets the **NSTC desirable characteristics**

Princeton Data Commons: Describe



New Submission

Describe | Readme | Upload | Review | Completed

Required Metadata

Additional Metadata

Curator Controlled

By initiating this new submission, we have reserved a draft DOI for your use. This DOI will not go live until the data review process is complete, but you can provide this DOI to a publisher now. This unique identifier will become searchable at the completion of this process.

10.80021/epa3-1520

Title* ?

Test Data Set

Description* ?

Please enter a brief summary describing the unique attributes of this submission. (e.g., The purpose, scope, methods, dependencies, etc. for the object(s) deposited to this submission.) It should be unique from a description or abstract used for a corresponding paper.

Rights* ?

Creative Commons Attribution 4.0 International

Creator(s)* ?

ORCID	Given name	Family name
0000-0001-9489-0750	Wind	Cowles

Add Another Creator | Add me as a Creator

Cancel | Save Work

Princeton Data Commons: Describe

Large-Eddy Simulation Results for Satellite-Sensed Sea Ice Surfaces

Creators: Fogarty, Joseph from Princeton University ROR ; Bou-Zeid, Elie from Princeton University ROR

Curator: Sarah E Reiff Conell

DOI: 10.34770/2ncq-sq05 COPY

Depositor: jf38

Description: Large-eddy simulations were employed over eight different satellite-sensed sea ice patterns to analyze how the overlying atmospheric boundary layer (ABL) dynamics and thermodynamics differ based on the heterogeneity of the surface. In Suite 1, four simulations were conducted for three different sea ice patterns, resulting in a total of 12 simulations. In Suite 2, two simulations were conducted for eight different sea ice patterns, resulting in a total of 16 simulations. Mean time-averaged variables presented include the geostrophic wind directions u, v, and w, potential temperature theta, and heat and momentum fluxes.

Keywords: simulations | sea ice pattern | atmospheric boundary layer

Domains: Natural Sciences

Rights: Creative Commons Attribution 4.0 International (CC BY)

Publisher: Princeton University

Publication Year: 2024

Version: 1

Resource Type: Dataset

General Type: Dataset

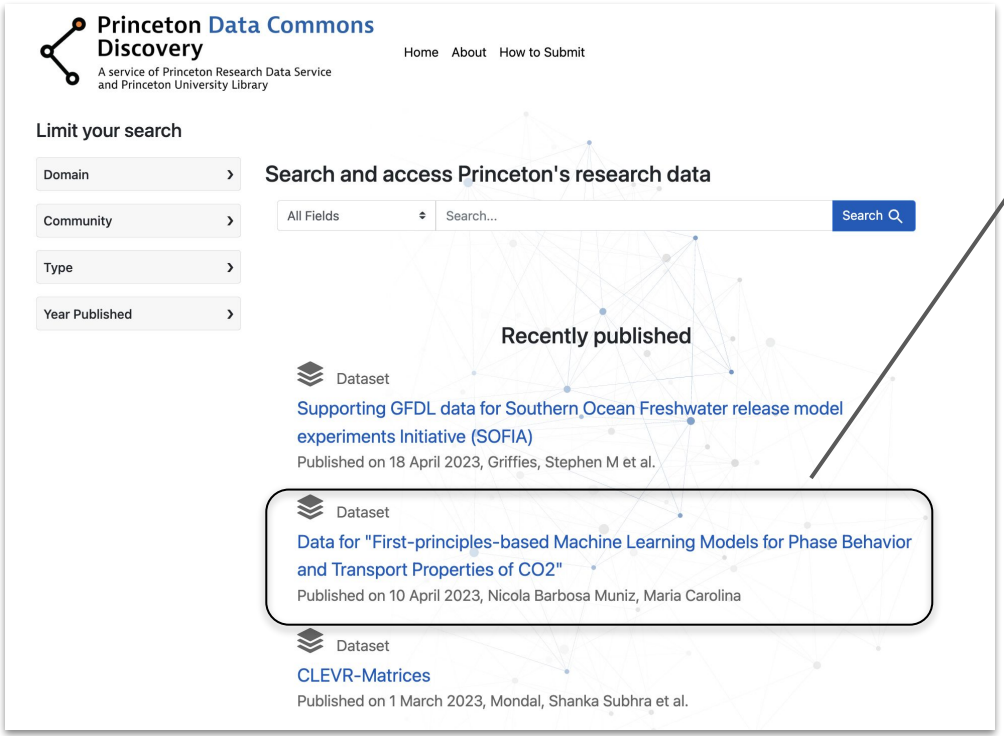
Discovery URL: https://datacommons.princeton.edu/discovery/catalog/doi-10-34770-2ncq-sq05

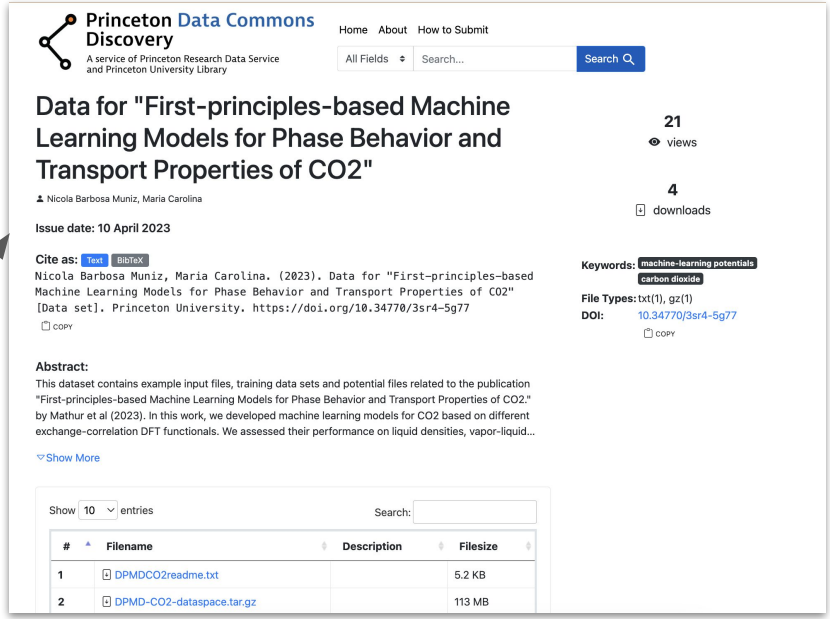
Group: Princeton Research Data Service (PRDS)

Funders: National Science Foundation, AGS 2128345,
National Oceanic and Atmospheric Administration, NA18OAR4320123,

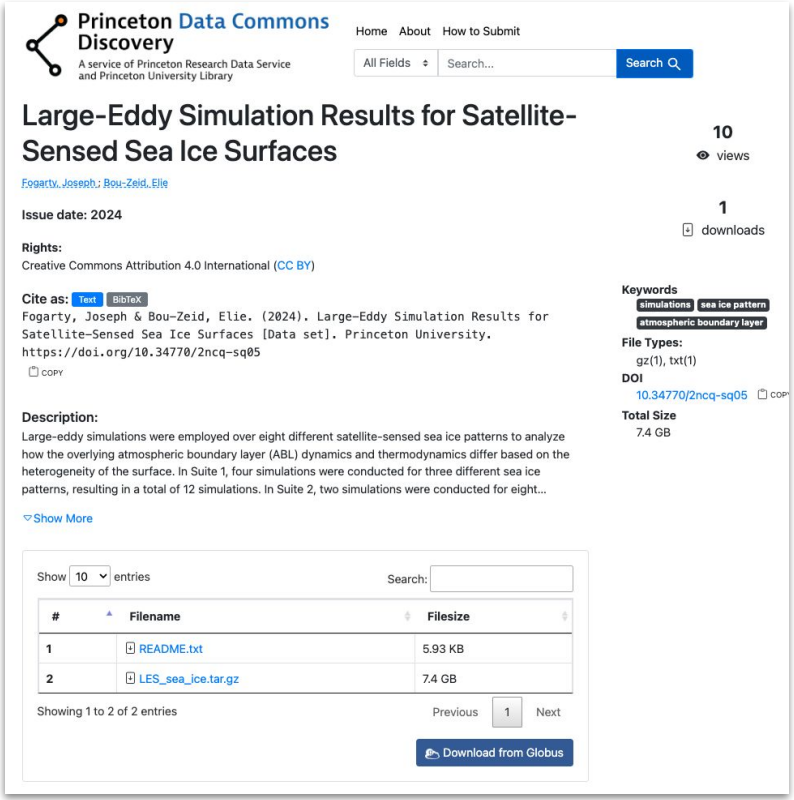
Location: Amazon S3 Curation Storage

Princeton Data Commons: Discovery

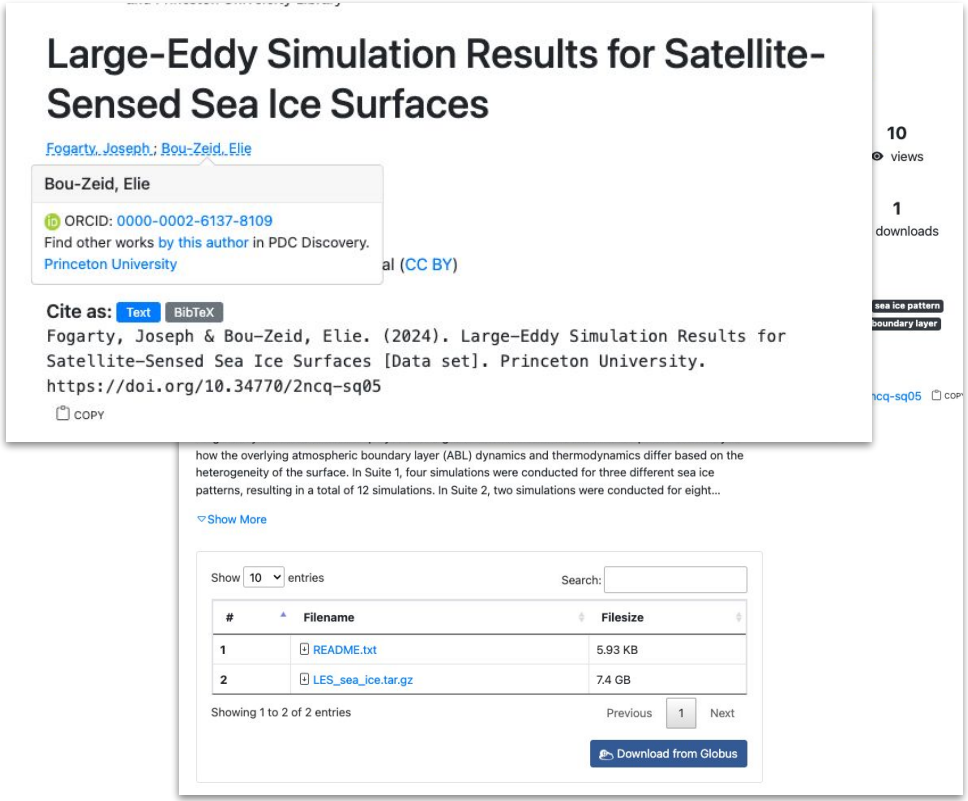




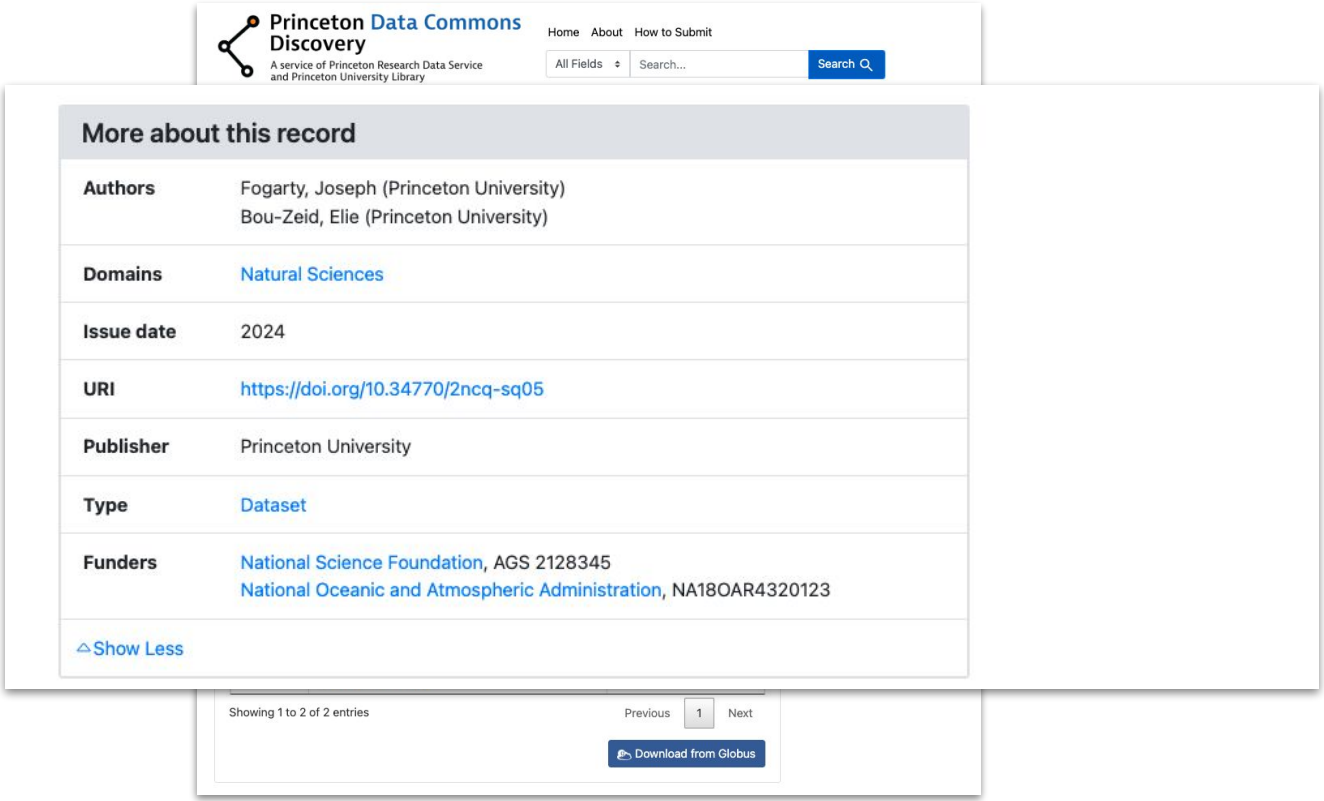
Princeton Data Commons: Discovery



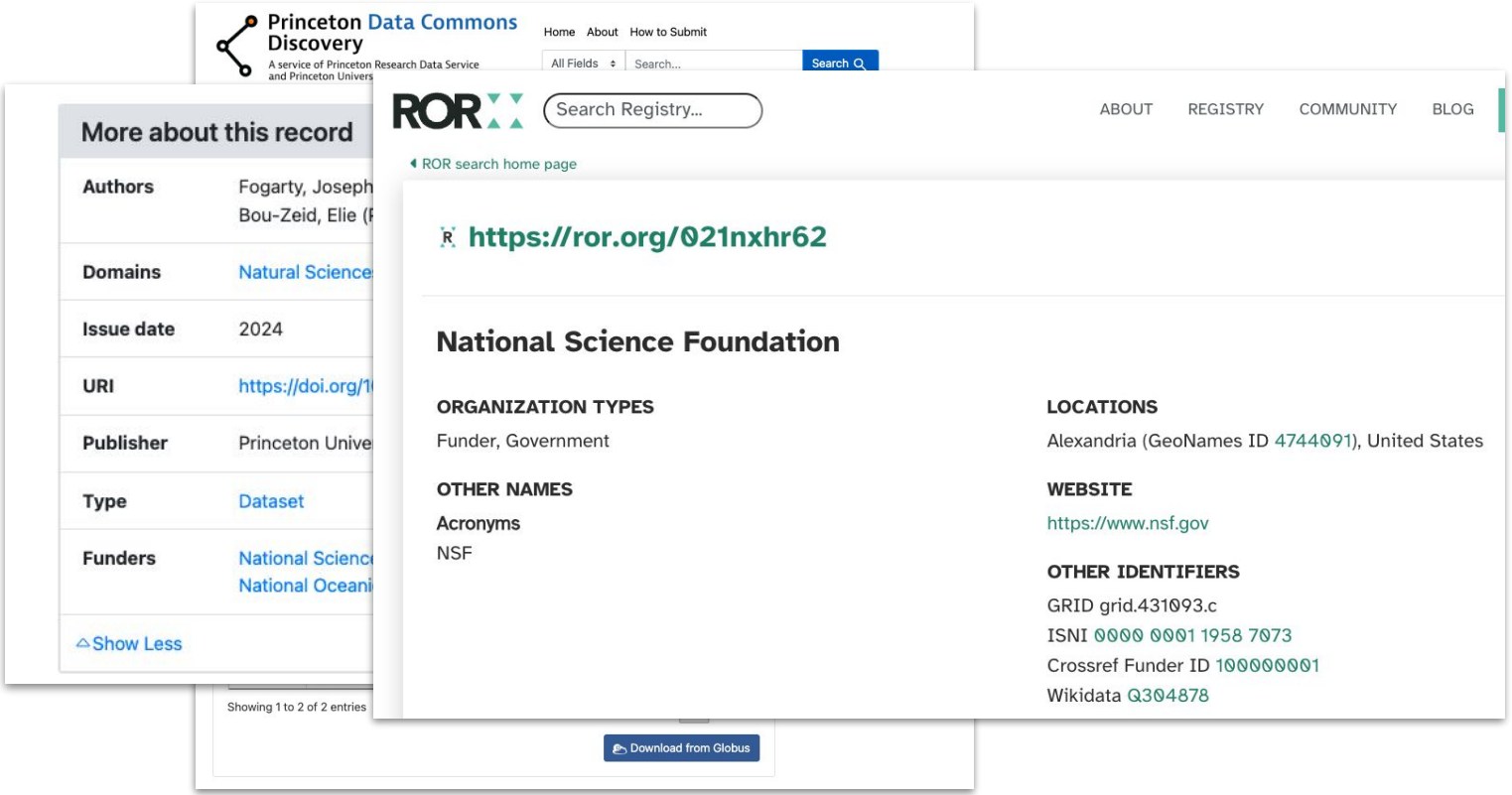
Princeton Data Commons: Discovery



Princeton Data Commons: Discovery



Princeton Data Commons: Discovery



Data Migration Process

- **Redescription process**
- **No automated workflows for data loading**

Data Migration Process

- **Redescription process**
- **No automated workflows for data loading**

- **Pros:**
  - Leverage domain expertise of users redescribing content from legacy system into new system
  - Exercise the new system
  - Address long-standing data issues
  - Detect work-stopping bugs in the new system faster
  - Learn how users **want** to use the system

- **Cons:**
  - Most feasible for a smaller amount of data and/or a large base of users
  - More time-consuming than automated workflows
  - Sunsetting legacy system takes more time

Challenges

Reorganization: new staff, new leadership, new roles

Stakeholders: Variety of administrative stakeholders across numerous campus and library units

Communications: Inherited communication structure across teams that no longer worked

Translating across teams: Two teams with different functional specialties, work culture, and work language

Competing priorities: this isn't the only application RDSS is building!

Opportunities

Reorganization: improved alignment of responsibilities, new key staff members, assignment of Product Owner role

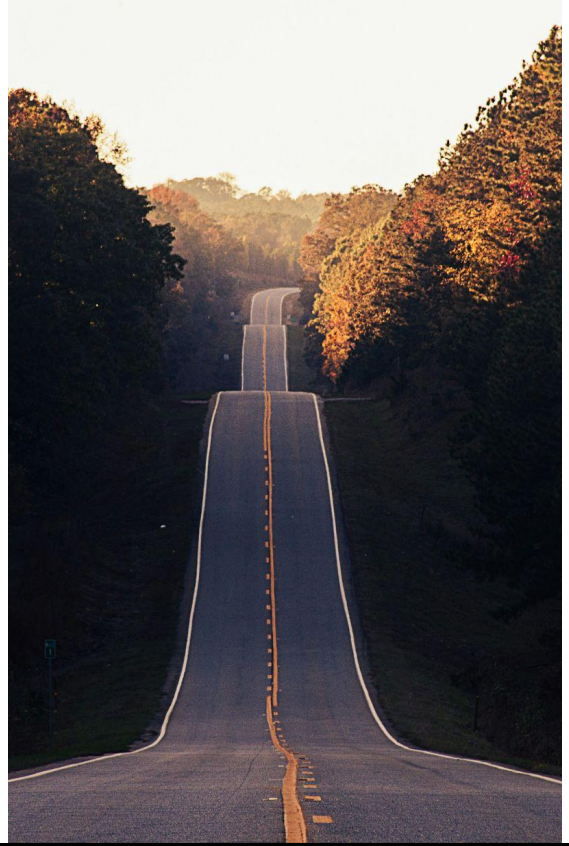
Stakeholders: Opportunity to reconvene stakeholders

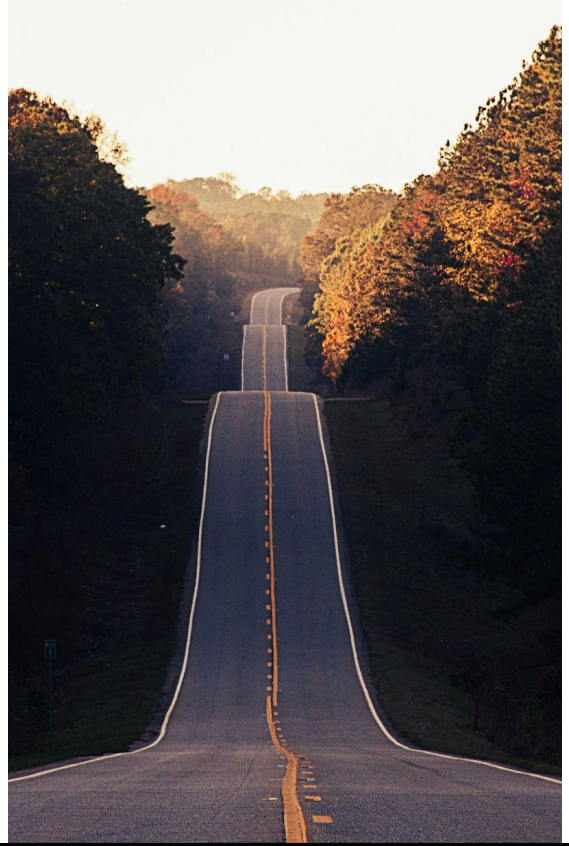
Communications: Opportunity to reassess and consolidate communication channels

Redefining priorities: Adopting a more realistic timeline

Conclusion

- **Timeline**
  - April 2024: New data submissions in Princeton Data Commons **only**
  - Summer 2024: Complete data redescription process
  - Fall 2024: Public launch and publicity
- **Product owner in PRDS**
  - New feature design and negotiation between IT and stakeholders



Conclusion

- **Thank you!**
  - katelynch@princeton.edu
  - mtesterman@princeton.edu