

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

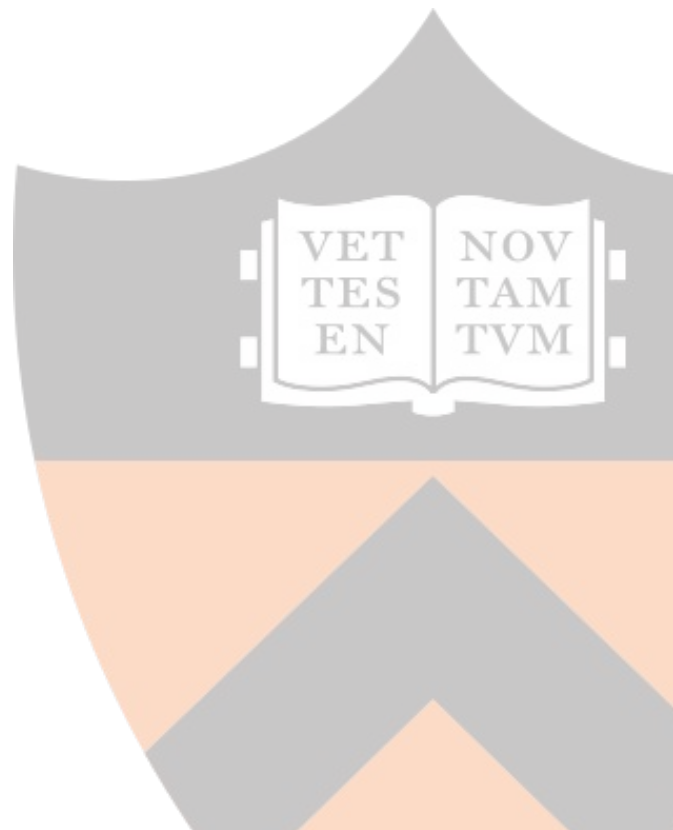
Meghan Testerman (she/her)

Open Research and Scholarship Librarian

Head, Princeton Research Data Service

Princeton University

mtesterman@princeton.edu



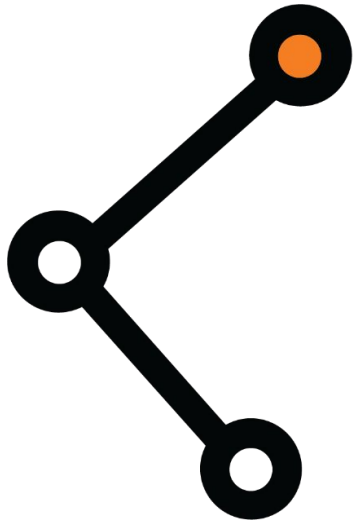
Introductions



Princeton University Library



Princeton **Research
Data Service**



Princeton Data Commons

How we got here - DataSpace



Home DataSpace About Contact Help Search Explore - Login -

DataSpace at Princeton University

Princeton's Research Data Repository is becoming the Princeton Data Commons! As we transition, you may not see some of the collections you were once able to submit to in DataSpace, and that means Princeton Data Commons is ready for you. If you are intending to publish from the Princeton Plasma Physics Laboratory and you have questions about the new submission process, please contact publications@pppl.gov. For all other questions about getting started with submissions to Princeton Data Commons, please contact prds@princeton.edu. For more information, please visit the [Princeton Research Data Service website](#).

DataSpace is a digital repository meant for both archiving and publicly disseminating digital data which are the result of research, academic, or administrative work performed by members of the Princeton University community. DataSpace will promote awareness of the data and address concerns for ensuring the long-term availability of data in the repository.

Communities in DataSpace

Astrophysical Sciences	Liechtenstein Institute on Self-Determination
Chemical and Biological Engineering	Mechanical and Aerospace Engineering
Chemistry	Molecular Biology
Civil and Environmental Engineering	Music and Arts
Computational Social Science	Office of Information Technology
Department of Geosciences	Physics
Department of Slavic Languages and Literatures	Princeton Neuroscience Institute
Digital Humanities	Princeton Plasma Physics Laboratory
Economics	Princeton School of Public and International Affairs
Education Research Section	Princeton University Doctoral Dissertations, 2011-2024
Electrical and Computer Engineering	Princeton University Library
Faculty Publications	Princeton University Masters Theses, 2022-2024
Geophysical Fluid Dynamics Laboratory	Princeton University Undergraduate Senior Theses, 1924-2023
Industrial Relations Section	Psychology
Lewis-Sigler Institute for Integrative Genomics	Seeger Center for Hellenic Studies
	Sociology

How we got here - DataSpace



The screenshot shows the DataSpace website interface. At the top, there is a navigation bar with 'DataSpace', 'About', 'Contact', and 'Help'. A search bar is on the right with 'Search', 'Explore', and 'Login' options. The main content area is titled 'Chemical and Biological Engineering / Research Data Sets'. It features a metadata section with fields for Title, Contributors, Keywords, Issue Date, Publisher, Abstract, URI, Referenced By, and Appears in Collections. A 'Files in This Item' table lists two files: 'DPMDCO2readme.txt' (5.2 kB, Text) and 'DPMDCO2-dataspace.tar.gz' (115.66 MB, Unknown). A 'Show full item record' button is at the bottom left, and a copyright notice is at the bottom right.

[DataSpace](#) [About](#) [Contact](#) [Help](#) [Explore](#) [Login](#)

[Home](#) / [Chemical and Biological Engineering](#) / [Research Data Sets](#)

Please use this identifier to cite or link to this item: <http://arks.princeton.edu/ark:/88435/dsp01d52w794k>

Title: Data for "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO₂"

Contributors: Mathur, Reha
Yue, Shuwen
Car, Roberto
Panagiotopoulos, Athanasios
Nicola Barbosa Muniz, Maria Carolina

Keywords: machine-learning potentials
carbon dioxide

Issue Date: 10-Apr-2023

Publisher: Princeton University

Abstract: This dataset contains example input files, training data sets and potential files related to the publication "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO₂" by Mathur et al (2023). In this work, we developed machine learning models for CO₂ based on different exchange-correlation DFT functionals. We assessed their performance on liquid densities, vapor-liquid equilibrium and transport properties.

URI: <https://doi.org/10.34770/3sr4-5g77>
<http://arks.princeton.edu/ark:/88435/dsp01d52w794k>

Referenced By: R. Mathur, M. C. Muniz, S. Yue, R. Car and A. Z. Panagiotopoulos, First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO₂. Submitted to Journal of Physical Chemistry B. (2023) Link to the article will be added upon publication.

Appears in Collections: [Research Data Sets](#)

Files in This Item:

File	Description	Size	Format	
DPMDCO2readme.txt		5.2 kB	Text	View/Download
DPMDCO2-dataspace.tar.gz		115.66 MB	Unknown	View/Download

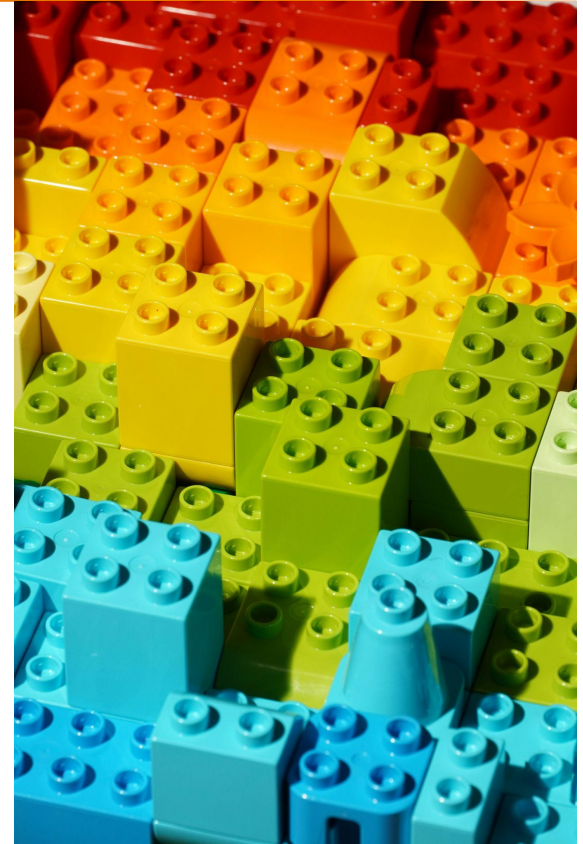
[Show full item record](#)

Items in DataSpace are protected by copyright, with all rights reserved, unless otherwise indicated.

The solution



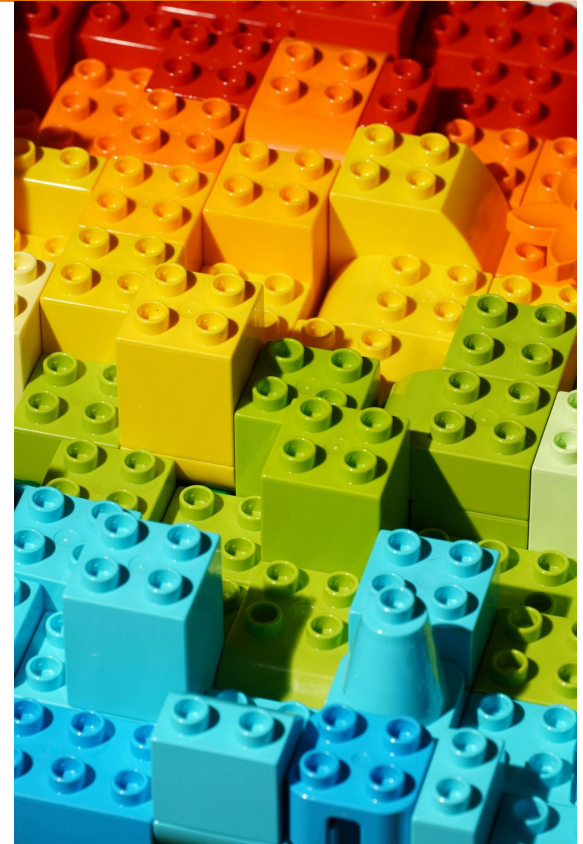
- **Princeton Data Commons (PDC)**
 - Blacklight and Ruby on Rails
 - Code repositories
 - https://github.com/pulibrary/pdc_discovery
 - https://github.com/pulibrary/pdc_describe

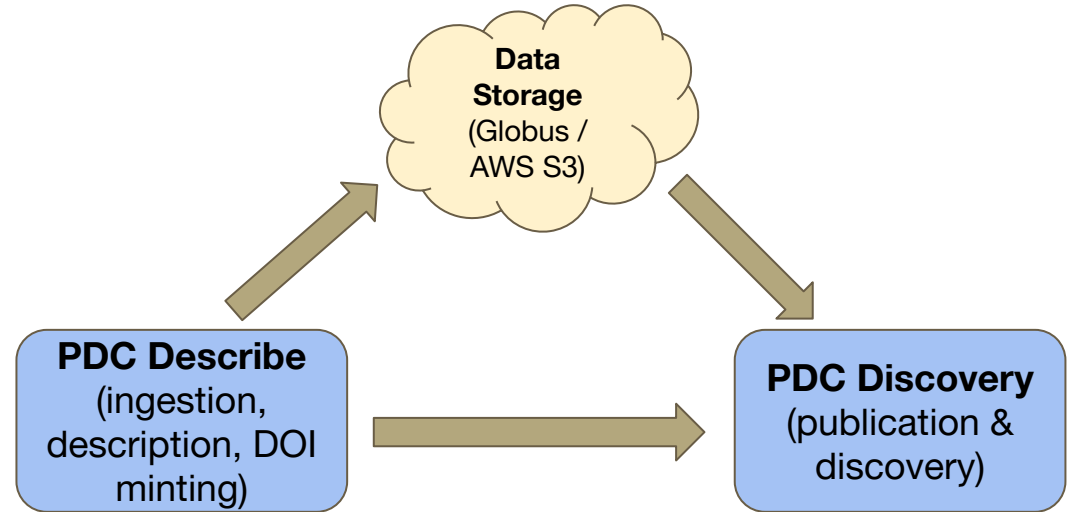
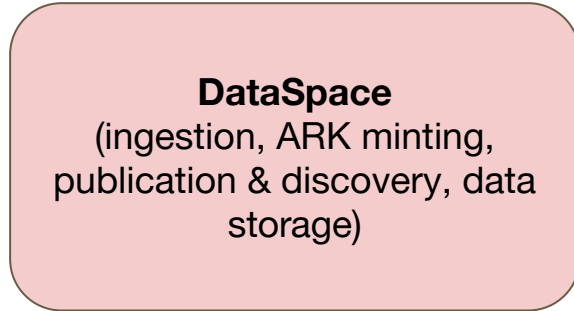


The solution



- **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 - 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**



New Submission

Describe Readme Upload Review Completed

You are starting a new submission. By clicking on the "Create New" button, a draft DOI will be reserved for this deposit.

If you have already started a submission for this deposit, then please [go your dashboard](#) and complete the existing submission. You can find more documentation on the [PRDS Data Repository Help Page](#). If you would like to talk to a specialist about your submission, we may be reached at prds@princeton.edu

Title* [?](#)

Creator(s)* [?](#)

ORCID

Given name

Family name



Add Another Creator

Add me as a Creator

Go Back

Create New

New Submission

Describe Readme Upload Review Completed

Required Metadata

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.

Additional Metadata

10.80021/epa3-1520

Curator Controlled

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

0000-0001-9489-0750

Given name

Wind

Family name

Cowles



Add Another Creator

Add me as a Creator

Cancel

Save Work

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:

DOI: [10.34770/2ncq-sq05](https://doi.org/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 θ , 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
A service of Princeton Research Data Service and Princeton University Library

Home About How to Submit

Limit your search

- Domain >
- Community >
- Type >
- Year Published >

Search and access Princeton's research data

All Fields Search...

Recently published

Dataset
[Supporting GFDL data for Southern Ocean Freshwater release model experiments Initiative \(SOFIA\)](#)
Published on 18 April 2023, Griffies, Stephen M et al.

Dataset
[Data for "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO2"](#)
Published on 10 April 2023, Nicola Barbosa Muniz, Maria Carolina

Dataset
[CLEVR-Matrices](#)
Published on 1 March 2023, Mondal, Shanka Subhra et al.

Princeton Data Commons Discovery
A service of Princeton Research Data Service and Princeton University Library

Home About How to Submit

All Fields Search... Search Q

Data for "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO2"

Nicola Barbosa Muniz, Maria Carolina

Issue date: 10 April 2023

Cite as: [text](#) [BibTeX](#)
Nicola Barbosa Muniz, Maria Carolina. (2023). Data for "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO2" [Data set]. Princeton University. <https://doi.org/10.34770/3sr4-5g77>

21 views

4 downloads

Keywords: [machine-learning potentials](#) [carbon dioxide](#)

File Types: [txt\(1\)](#), [gz\(1\)](#)
DOI: [10.34770/3sr4-5g77](https://doi.org/10.34770/3sr4-5g77)

Abstract:
This dataset contains example input files, training data sets and potential files related to the publication "First-principles-based Machine Learning Models for Phase Behavior and Transport Properties of CO2." by Mathur et al (2023). In this work, we developed machine learning models for CO2 based on different exchange-correlation DFT functionals. We assessed their performance on liquid densities, vapor-liquid...

Show More

Show 10 entries

#	Filename	Description	Filesize
1	DPMDCO2readme.txt		5.2 KB
2	DPMDCO2-dataspaces.tar.gz		113 MB

Princeton Data Commons: Discovery



Princeton Data Commons Discovery
A service of Princeton Research Data Service and Princeton University Library

Home About How to Submit

All Fields Search... Search

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

Fogarty, Joseph; Bou-Zeid, Elie

Issue date: 2024

Rights: Creative Commons Attribution 4.0 International (CC BY)

Cite as: [Text](#) [BibTeX](#)
Fogarty, Joseph & Bou-Zeid, Elie. (2024). Large-Eddy Simulation Results for Satellite-Sensed Sea Ice Surfaces [Data set]. Princeton University.
<https://doi.org/10.34770/2ncq-sq05>
[copy](#)

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...

[Show More](#)

10 views
1 download

Keywords
[simulations](#) [sea ice pattern](#)
[atmospheric boundary layer](#)

File Types:
gz(1), txt(1)

DOI
[10.34770/2ncq-sq05](https://doi.org/10.34770/2ncq-sq05) [copy](#)

Total Size
7.4 GB

Show 10 entries Search:

#	Filename	Filesize
1	README.txt	5.93 KB
2	LES_sea_ice.tar.gz	7.4 GB

Showing 1 to 2 of 2 entries Previous 1 Next

[Download from Globus](#)

Princeton Data Commons: Discovery



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

[Fogarty, Joseph](#); [Bou-Zeid, Elie](#)

Bou-Zeid, Elie

ORCID: [0000-0002-6137-8109](#)

Find other works by this author in [PDC Discovery](#).
[Princeton University](#)

al (CC BY)

Cite as: [Text](#) [BibTeX](#)

Fogarty, Joseph & Bou-Zeid, Elie. (2024). Large-Eddy Simulation Results for Satellite-Sensed Sea Ice Surfaces [Data set]. Princeton University.
<https://doi.org/10.34770/2ncq-sq05>

COPY

10

views

1

downloads

sea ice pattern
boundary layer

ncq-sq05 COPY

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...

[Show More](#)

Show entries

Search:

#	Filename	Filesize
1	README.txt	5.93 KB
2	LES_sea_ice.tar.gz	7.4 GB

Showing 1 to 2 of 2 entries

Previous Next

Download from Globus

Princeton Data Commons: Discovery



Princeton Data Commons Discovery
A service of Princeton Research Data Service and Princeton University Library

Home About How to Submit

All Fields Search... Search Q

More about this record

Authors	Fogarty, Joseph (Princeton University) Bou-Zeid, Elie (Princeton University)
Domains	Natural Sciences
Issue date	2024
URI	https://doi.org/10.34770/2ncq-sq05
Publisher	Princeton University
Type	Dataset
Funders	National Science Foundation , AGS 2128345 National Oceanic and Atmospheric Administration , NA18OAR4320123

[Show Less](#)

Showing 1 to 2 of 2 entries

Previous 1 Next

[Download from Globus](#)

Princeton Data Commons: Discovery



The screenshot displays the Princeton Data Commons Discovery interface. At the top, the header includes the Princeton Data Commons logo, navigation links (Home, About, How to Submit), and a search bar. Below the header, a sidebar on the left titled "More about this record" lists metadata: Authors (Fogarty, Joseph; Bou-Zeid, Elie), Domains (Natural Science), Issue date (2024), URI (https://doi.org/10.21203/rs.3.rs-3811111/v1), Publisher (Princeton University), Type (Dataset), and Funders (National Science Foundation, National Oceanic and Atmospheric Administration). The main content area shows the ROR search results for the record, displaying the ROR ID <https://ror.org/021nxhr62> and the organization name "National Science Foundation". Below this, the record details are organized into sections: ORGANIZATION TYPES (Funder, Government), OTHER NAMES (Acronyms: NSF), LOCATIONS (Alexandria (GeoNames ID 4744091), United States), WEBSITE (https://www.nsf.gov), and OTHER IDENTIFIERS (GRID grid.431093.c, ISNI 0000 0001 1958 7073, Crossref Funder ID 100000001, Wikidata Q304878). A "Download from Globus" button is visible at the bottom of the record details.

Princeton Data Commons Discovery
A service of Princeton Research Data Service and Princeton University

Home About How to Submit

All Fields Search...

ROR Search Registry...

ABOUT REGISTRY COMMUNITY BLOG

◀ ROR search home page

<https://ror.org/021nxhr62>

National Science Foundation

ORGANIZATION TYPES
Funder, Government

OTHER NAMES
Acronyms
NSF

LOCATIONS
Alexandria (GeoNames ID 4744091), United States

WEBSITE
<https://www.nsf.gov>

OTHER IDENTIFIERS
GRID grid.431093.c
ISNI 0000 0001 1958 7073
Crossref Funder ID 100000001
Wikidata Q304878

Showing 1 to 2 of 2 entries

Download from Globus

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 redescrbing 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

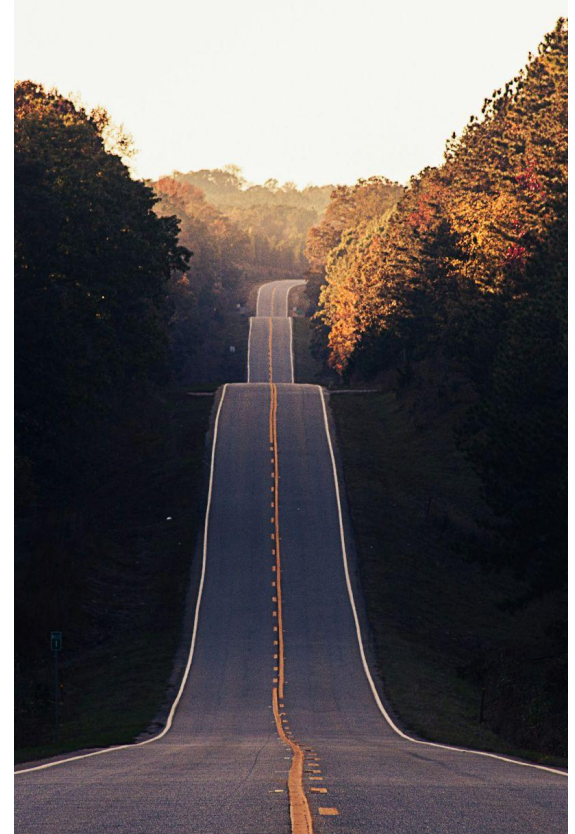
Communications: Opportunity to reassess and consolidate communication channels

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

