Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1W96D

Title: Data Transfer Service

Creator: Elisha Wood-charlson - **ORCID:** 0000-0001-9557-7715

Affiliation: Lawrence Berkeley National Laboratory (lbl.gov)

Principal Investigator: Elisha Wood-Charlson

Project Administrator: AJ Ireland

Contributor: Kjiersten Fagnan, Jeffrey N Johnson

Funder: United States Department of Energy (DOE) (energy.gov)

Template: Department of Energy (DOE): Office of Science (Updated 2022)

Project abstract:

The transfer of data between BER funded programs often requires researchers to manually download and upload files. This takes time, adds the potential for error (both human and data integrity), and often removes any citation/credit information for the data being transferred. We will build a data transfer service that can be used by any BER funded program to directly request or send data objects from other programs that is accompanied by appropriate provenance and citation metadata. By ensuring data object integrity and retention of credit metadata, DTS will enable better tracking of data reuse across the BER funded portfolio. Finally, the DTS aims to reduce the time and effort required by researchers to accomplish their science.

Start date: 10-01-2023

End date: 09-30-2025

Last modified: 01-22-2024

Copyright information:

language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Data Transfer Service

The Data Transfer Service will not generate any new data. The goal of the DTS is to more efficiently transfer existing data between data platforms, ensuring data integrity and credit metadata associated with the data object are intact and travel with the data itself.

DTS will connect BER data platforms, and any trusted non-BER data platform, that is interested in connecting to the service.

All software and data models developed for the DTS will be open source, and made available with robust documentat on GitHub. Documentation includes how to use the API, detailed information around fields in the credit metadata schema, and comprehensive tutorials on how to connect to the service via each participating data platform.

All code and transfer logs (source/destination, data identifier, md5) will be made available via the GitHub repository.

Question not answered.

DTS has requested funding to ensure that the production service is deployed on hardware that serves the JGI and KBase programs, in the Integrated Genomics Building on the Berkeley Lab campus.

Question not answered.

After the completion of this project, maintenance of the DTS will be performed by JGI and KBase team members involved in ongoing JGI-KBase co-development efforts.

Question not answered.

No PII will be stored by the DTS. Authentication protocols will be used to ensure transfer of data between portals (e.g., ORCID). Once the transfer is complete, authentication information is not retained. Transfer logs are only to track the location of data across the portals, to avoid duplicate transfers and to quantify movement of data between programs.

The DTS is not generating new data, but will greatly improve the researcher experience in needing to move data from one or more portals. The aim is to simplifying curation efforts, streamlining the steps between data production and publication, and directly support FAIR and open data best practices.

Planned Research Outputs

Software - "Data Transfer Service"

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Data Transfer Service	Software	Unspecified	Open	None specified		Creative Commons Zero v1.0 Universal	None specified	No	No