

A space-themed illustration on a dark blue background. On the left, a white and orange rocket with a blue nose cone is launching from a white launch pad. In the center, a white moon with blue spots is visible. To the right, a large orange planet with horizontal stripes represents Mars. The scene is decorated with several yellow stars and two shooting stars with white trails.

ACQUISITION OF AND ACCESS TO RESEARCH OMICS DATA

2024 Human Research Program
Investigators' Workshop

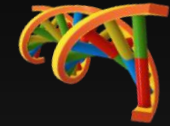
Dan Berrios, Greg Eley, Ashley Mooney, Deepak Kulkarni



Omic Data

- “Omic” originally derived from “genomic”, but now includes several kinds of comprehensive, “totality” measurement sets
 - Whole genome sequencings (DNA)
 - RNA transcript sequencing, messenger RNA: transcriptomics
 - All detectable metabolites: metabolomics
 - + 255 others: Metagenomics, proteomics, epitranscriptomics, etc.
- Variety of measurement types and formats
- Frequently large data volumes
- (Commonly) several levels of data processing yielding...
- ...tens to hundreds of physical files in a complex logical organization

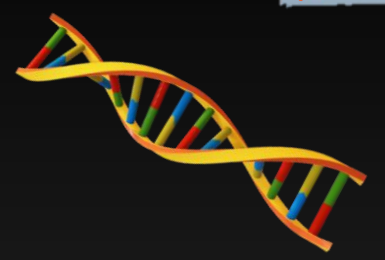
NASA Policy Directive for Genetic Data



- NPD 7170.1 governs storage and access to all “human genetic testing data” gathered for research purposes
 - Applicable to identifiable data (does not include aggregated or otherwise non-identifiable data)
 - Storage separated from Electronic Medical Record systems (unless otherwise authorized by subject)
 - Access to the data by certain personnel (e.g., NASA employment or duty managers) is prohibited
 - Storage of such data in NLSP would conform to this directive
- HRP DMP details the impacts of NPD 7170.1
 - “...Omics data are very broad in nature and may be abstracted and aggregated in a variety of ways which may remove any risk of identifiability. Additionally, omics data may represent information constitutional to a research subject, or may represent data that changes over time, based on environmental exposure. NPR 7170.1 places limitations on the control and use of omics data...”

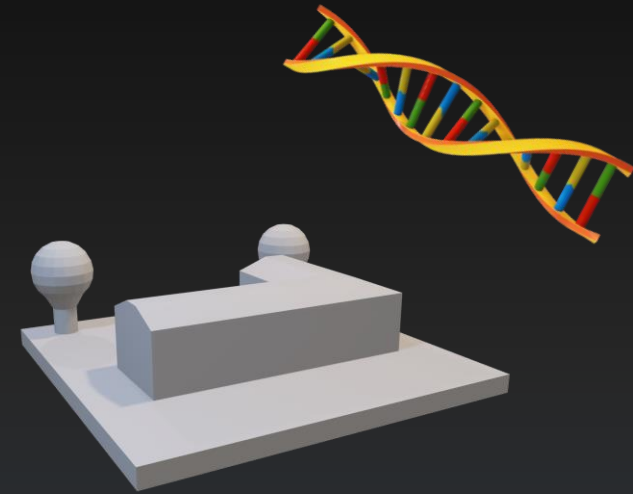


Data archiving

- DMP further specifies HRP Omics Data archival locations
 - *“Human omics data should reside within the LSDA, while non-human omics data may be directed to GeneLab.”*
 - *Investigators may opt to deposit omics data into additional repositories (certain research communities and/or publishers require such depositions)*
- 

Omic Data Archival Transfer

- Sources: Data systems used by
 - HRP Funded Investigator
 - HRP-contracted Facilities
 - Investigator-contracted Facility
- Destination:
 - NLSP Cloud Data Storage (currently AWS S3)
- Design Concerns:
 - Security, Reliability, Accessibility, Flexibility
- Submitters choose from a configurable list of transfer mechanisms





Omic Data Archival Transfer Options

Data Transfer Option	Maximum Single File Size	NLSP Release
Web Browser Upload	5 TB	3.4
Integrated Large File (Box) Transfer	150 GB	3.5
Integrated AWS S3 API Transfer	5 TB	4.0
AWS Snowball Device Transfer	20 TB (this will not fit in an AWS bucket)	4.5

Options for Omics Transfer: Web Browser

Web Browser Upload

Instructions

The table below lists the files currently uploaded to this submission package. To add files, click the "Upload New" icon below. You will then be able to upload files by clicking the "Browse" button or dragging and dropping the file. To edit metadata for a file, click the gear icon under the "Manage" column and select "Edit".

When all files have been submitted and all metadata is complete and accurate, click the box labelled "Mark as Complete" below the table. This will mark the File Upload step as complete. If needed, you may return to edit metadata or uncheck the completion box later. **However, once you send the submission package to the archivist for review, you cannot make changes to the submission package.**

Mark Uploaded Files As Complete



Drag and drop Files Here to Upload

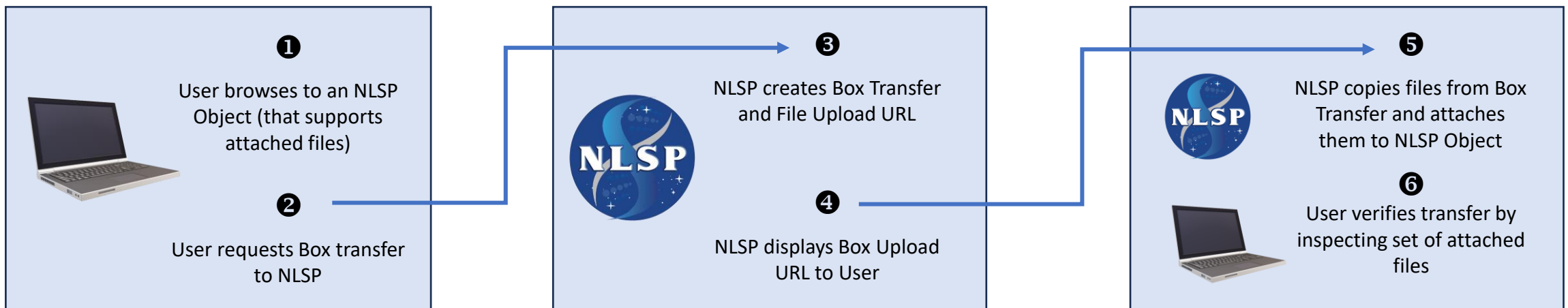
Or Select Files to Upload

Data Submission Files

0 records returned

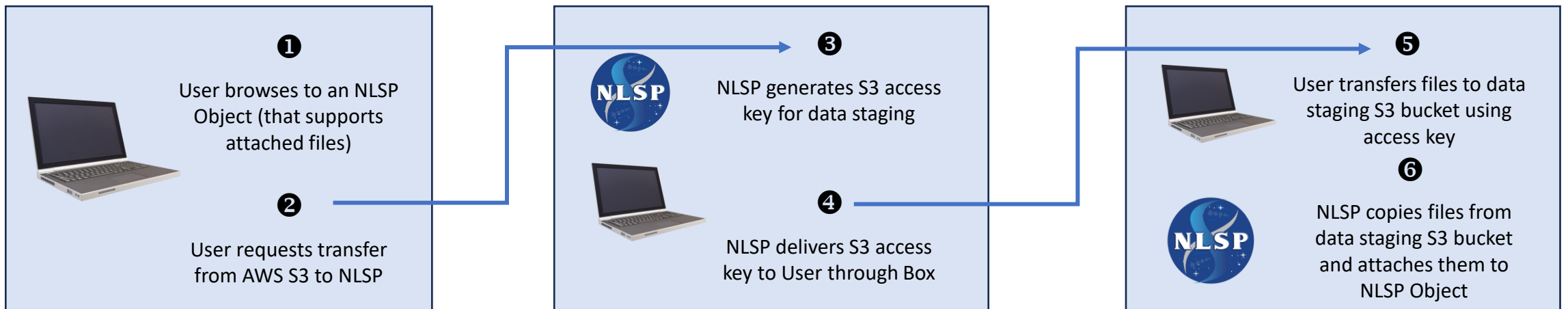
Options for Omics Transfer: Large File Transfer

Using NASA ATO Large File Transfer application (currently Box)




Options for Omics Transfer: AWS S3 API Transfer

Direct copying of files to NLSP AWS S3 data storage





Use of Controlled data in NLSP Workspaces

- It is anticipated that transferred (submitted) omics data will be a key target in the use of NLSP Workspaces
 - NLSP Release 4.0 to include Galaxy Analysis Environment
 - Configured to support Omics analyses
 - Access controls are in place within analytical systems, congruent with the NLSP Role Based Access Controls.
 - These can be configured to conform with NPD 7170.1
 - As specified by the HRP DMP
 - *“To support analysis of controlled data within the NLSP workspace, role-based access must be extended to the workspace to ensure sensitive data are not exposed to users lacking the appropriate authorizations.”*
- 



Summary

- The NLSP Data Submission Portal design supporting omics data transfer
 - Complies with NASA direction to restrict access to research genetic testing data
 - Leverages cloud-based storage for optimum data reliability and integrity
 - Provides data submitters with several options tailored to their skill level and available resources
 - Supports a wide range of transfer volumes and complexities
 - Is extensible by design to adapt existing or include additional transfer options when they become available
 - Enables data validation via integrated workspaces



Acknowledgements

- NLSP: Carlos De Los Santos, John Dunn, Truong Le, Sandeep Shetye, Patrick Moore
- DMIO: Olga Kieschnick
- LSDA: Jessica Keune
- Others: Frank King, Eric Gallagher, Milt Checchi, Dawayne Pretlor, Andre Dizon, Matthew Rodriguez, Lai Nguyen, Brenna Proulx, Daniel Dimarco