The Data Conservancy Packaging Tools and Use Perspectives from Research Data Management Services

Tim DiLauro and Jonathan Petters | Johns Hopkins University Sheridan Libraries | datamanagement@jhu.edu

Research Data Services Needs
• N1) To reduce complexity of research data transfer from researchers to the data archive.
• Ability to capture N2) relationships between and N3) fixity information for research data objects, and do both as close as possible to researchers.
• N4) Ability of the data archive to augment data from researchers while ensuring original content is not lost or changed.

Challenging Use Cases in JHUDMS Implementation
• U1) Creating package on one system and augmenting research data on another system, possibly with a different OS.
• U2) Using the Data Conservancy Packaging Tool from an external hard drive, connected to a researcher’s system.
• U3) Creating package and package description files with names that can conflict.

Data Conservancy Packaging Tools Capabilities
• C1) Creates single package file, avoiding OS-specific issues such as hidden files and misinterpreted characters during research data transfer.
• C2) Captures checksums for packaged research data objects, allowing for end-to-end verification of object integrity.
• C3) Captures graph, allowing for expression of relationships among packaged research data objects and external resources.

Future Enhancements to the Tools
• Ability to open and modify existing package descriptions and packages
• Generalizing information model through ontology support
• Additional serializations (e.g., Research Object Bundle format)

JHU Data Management Services

Special thanks to the Data Conservancy Infrastructure R&D Team