

Automation of Preservation Functions

REAGAN W. MOORE, RICHARD MARCIANO, ANTOINE DE TORCY, AND PAUL WATRY

Abstract: The NARA Transcontinental Persistent Archive Prototype uses policy-based data management systems to implement preservation environments. In collaboration with the SHAMAN project (Sustainable Heritage Access through Multivalent ArchiviNg), an end-to-end demonstration of a preservation environment is being constructed. The system integrates the Cheshire workflow system with the integrated Rule-Oriented Data System to automate accession, arrangement, description, and preservation of records. Specific functions include ingestion of Submission Information Packages, creation of Archival Information Packages, replication of records to minimize risk of data loss, periodic validation of checksums to verify integrity, parsing of records to extract descriptive and provenance metadata, and text based searching across both metadata and records. Validation functions include creation of log files to track all errors, parsing of log files to track compliance, and periodic execution of integrity and authenticity checks. We will discuss the use of policies to control these procedures, and provide examples of explicit rules used to validate the trustworthiness of the environment.

About the author:

Reagan Moore is a Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill, Chief Scientist for Data Intensive Cyber Environments at the Renaissance Computing Institute, and Director of the Data Intensive Cyber Environments Center at UNC. He coordinates research efforts in development of data grids, digital libraries, and preservation environments. Developed software systems include the Storage Resource Broker data grid and the integrated Rule-Oriented Data System. He has a Ph.D. in plasma physics from the University of California, San Diego, (1978) and a B.S. in physics from the California Institute of Technology (1967).