

Global Format Registry Use Case: Clarification of Data Format Version

Use Case ID	NYU-2.
Description	SIPs submitted to an OAIS-compliant repository may contain insufficient technical metadata to allow the data format for a packaged file to be completely identified. An agent for the repository uses the limited technical metadata to query the Registry for possible data formats which match the metadata, and representation information allowing the files conforming to the different formats to be differentiated. This information is then used to test the packaged data file to see if it conforms to a known data format recorded in the Registry.
Actors	Registry – accepts queries for data formats which may match limited technical metadata and returns representation information for those formats Repository agent – examines technical metadata to determine whether it precisely identifies a data type for a file, and if not, queries the Registry for potential data types based on metadata available.
Assumptions	Registry must have legal authority to disseminate representation information regarding format to requesting agent. Repository Agent must possess technical information regarding data file within a SIP to formulate a valid search request to send to the Registry. Repository Agent must be able to use returned representation information to confirm format of data file
Preconditions	Registry supports search mechanisms (keyword, probabilistic matching?) allowing it to identify potential matching data formats based on limited information. Registry and Repository agent must share application protocol for request/dissemination of representation information.
Triggers	Arrival of Submission Information Package which does not completely identify data format(s) for packaged file(s).
Primary Scenario	Step 1 -- Repository agent examines SIP to see if it uses formal identifiers (from Registry) to identify data formats for packaged content files. If so, then see NYU-1 use case. If not, then.... Step 2 – Repository agent examines SIP for any technical metadata which might assist in identifying data format for packaged files, formulates a query to the Registry based on that metadata, and submits query to Registry. Step 3 – Registry executes query and returns representation information for data formats which match query to Repository agent. Step 4 – Repository agent uses representation information to analyse packaged data files and determine if they conform to a known type.
Primary Result	Data files lacking proper identification of their data format within a SIP are properly identified.
Post-Conditions	
Non-functional requirements	
Notes	

Issues	
--------	--