Transport Neutral Digital Object Replication
Mark Phillips, Kurt Nordstrom; University of North Texas; Denton, Texas, USA

**IDEA**

The University of North Texas (UNT) Libraries has implemented a simple transport neutral digital object replication strategy in its production digital repository infrastructure. This strategy is built with the same ideals as other Curation Micro-Services, in respect to lightweight, software independent specifications coupled to provide a set of services for digital repositories. Building on standard Web technologies and methodologies like the Atom Publishing Protocol and REST, coupled with digital library technologies such as Checkm and BagIt, a transport neutral replication strategy allows institutions to meet the increasing demands on their services while keeping the overall costs low by allowing the use of a variety of storage platforms.

**THEORY**

1. **WHAT**
   - Objects do we want to replicate?

2. **WHERE**
   - Where can we get the contents of an object?

3. **HOW**
   - How do we verify the object once we have it?

- All Identifiers
- Manifest

**GOAL**

1. The source Digital Object Archive provides a listing of identifiers for contained objects.
2. For each identifier within the Digital Object Archive, we can request a listing of URLs for individual files.
3. Objects are packaged according to the BagIt specification, allowing for validity checking via contained manifest files.

- URL Listing
- Transport Neutral Digital Object Replication
- Mark Phillips, Kurt Nordstrom; University of North Texas; Denton, Texas, USA

**PRACTICE**

This diagram illustrates the actual steps involved in our replication process.

- Populator
- Replication Queue
- Backup Archive
- Master Archive
- Harvester
- Secondary Archive
- Master Archive
- Tertiary Archive
- HTTP Transport
- iRODS Transport
- "Not one possibility among many"

The populator queries the Master and Backup archives for content listings, and creates a queue of objects to replicate in the the Replication Queue.

The Harvester reads items from the Replication Queue, downloads them from the Master Archive and then stores them in the Backup Archive, checking them for validity before moving them to their final archival destinations and removing the item from the queue.

The goal is to provide seamless replication across a variety of storage and transport mediums, provided that each system is able to provide the necessary services for its contained objects.

- Replication Queue
- Master Archive
- Backup Archive
- Tertiary Archive
- HTTP Transport
- iRODS Transport
- "Not one possibility among many"