Name: Report Format Risk and Possible Migration Paths for a Known Resource

Author: Tim DiLauro

Goal: Determine if a known resource or any of its dependencies are at risk for format obsolescence and, if so, produce a report of potential migration paths for the affected resources.

Actors: Requesting Agent (human or system), Auditor (system), Format Registry (system), Preservation Planner (generally a human)

Summary: Over time, file formats become obsolete or unavailable due to changes in available technical environments. In order to avoid loss of access to needed content in these formats, the onset of these events must be detected and strategies developed to migrate this content to formats in supported technical environments.

Pre Conditions:
- The resource, its descendents, and their dependencies have PIDs and the PIDs have associated data describing structure (resource type and, for bytestreams, the file format).

Triggers:
- Requesting Agent determines that a resource should be evaluated for obsolescence.

Normal Path:
1. The Requesting Agent contacts the Auditor and requests an obsolescence audit, providing the PID of the target resource.
2. The Auditor uses the PID to access the resource’s PID metadata.
   a. If the resource is a bytestream, the Auditor adds its PID and format to a list of resources to review.
   b. The Auditor obtains a list of the PIDs of the resource’s descendents and their dependencies and recursively processes each of them via step #2.
3. For each unique format on the list, the Auditor queries the Format Registry to determine the obsolescence risk.
   a. If the risk is low enough, items matching that format are removed from the list.
   b. Items remaining on the list are at sufficient risk to warrant further review.
4. For each unique format in the remaining items, the Auditor requests a list of viable migration paths from the Format Registry.
5. The Auditor generates a summary list of at-risk formats with associated migration paths.
6. The Auditor generates a complete list of at-risk resources, their formats, and associated migration paths.
7. The Auditor forwards these lists to the Preservation Planner.
8. The Auditor notifies the Requesting Agent that the process has completed.

Alternate Paths:
- **2a.** The PID metadata record for some bytestream resources may not have format information.
- **2b.** The PIDs of some or all child or constituent resources are not captured in the PID metadata record: The Auditor would need to dereference the PID to extract the identifiers of these resources, if they are available.
- **3.** The Format Registry might not have obsolescence risk information for all formats. The Auditor would need to default to either removing those items from the list, leaving them on the list as-is, or leaving them on the list and noting that risk information was unavailable. This might be driven by some policy or heuristics.
- **4.** The Format Registry may not contain migration paths for some formats. The lists should indicate that there is no migration path for such formats.

Post Conditions:
- The Preservation Planner has a list of at-risk resources and some possible migration paths for some or all of them.

Notes:
- This case does not fully account for a conceptual resource like a Data Item (or Data Object), which might have multiple manifestations (representations) of the same information content as a result of previous format migrations/transformations. For such an object, the Auditor would need to determine if the resource has one or more manifestations that are not at risk, and if so, remove that resource from the list.