

Design - Fixity Service

User Stories and Features

(S = supported) or effort required to support = E (easy) M (medium) H (hard)

Priority = 1 must have, 2 should have, 3 nice to have

Story/Feature	Priority	RipRap	UMD	Camel Toolbox
Scheduling and Coordination				
Check all resources every n months (not appropriate for very large repos)		S		M
Run continual fixity checks		S		M
Only alert me when are there failures		S	S	E
Call an arbitrary HTTP endpoint on success and/or failure		S		E
Send a (jms,stomp) message to a endpoint on success and/or failure		M	S	E
Call an arbitrary HTTP endpoint on success and/or failure		M		E
Perform fixity check on most recent version of resource only		S	S	S
Perform fixity check on all versions of resource		S		S
Check entire repo		S	S	S
Deliver fixity requests in batches		S		M
Check resources based on query		S		M
Scaling				
Set rate limits (max bits/second) for checks		H		M
Set rate limits (max bits/second) for each task processor		H		M
Allow checks to be scaled horizontally across multiple instances		M		H
Run fixity against on disk representation		S		H
Reporting				
Generate list of resources checked since X date		S	S	M
Generate list of resources not checked X date		S	S	M
Generate a fixity audit report by resource		S	S	M
Generate a csv of fixity audit results showing resource id, date checked, checksum, and the result based on resource, date, and result parameters		S	S	M
Storage options				
Store fixity results in triplestore		M	S	
Store fixity results in text file		S		S
Store fixity results in sql		S		E
Store fixity results in nosql		M		E

Components

Fixity Checker

performs the fixity check on a resource and communicates the results to the configured Fixity Result Handler

Fixity Check Manager

- Is responsible for monitoring and providing status of all fixity check tasks
- coordinates parallel processing of fixity checks
- coordinates auto-scaling requests

Fixity Result Handler

A service responsible for storing the results. I can imagine different implementations that talk to different types of data storage backends

Fixity Result Reporter

A service that generates various fixity reports based on a various criteria.

Fixity Check Task Scheduler

A service responsible for generating Fixity Check requests across a repository based on user criteria such as min and max times between fixity checks, whether to run continuously or on a schedule, etc.