

Figure DB Schema

**DataSet table**

Language is the three letters ISO code (http://en.wikipedia.org/wiki/ISO\_3166-1\_alpha-3)

Type

Currently ESE is the only accepted metadata format but in the future we almost certain extend the harvesting to other formats such as LIDO. This field will contain the name of the metadata format.

**Provider table**

Country is two letter ISO code 3166-1-alpha2 (http://en.wikipedia.org/wiki/ISO\_3166-1\_alpha-2)

Type enumeration (for the moment)

Museum

Archive

Library

**Request Table**

This table identifies a specific harvesting for a given data set. It also indicates if this request can be sent to production.

During the loop where Repox is parsing the file the request should be in the state “under construction”, and Repox may abort the request and set the status accordingly, when this process is done the state must move to “import completed” and after this point Repox cannot change the data.

Request status

under construction - [r] repox is creating a new request

import completed - [r] repox ready, sip can take control when ready

aborted  - [rs] something went wrong

sip processing - [s] sip has found the request repox may not any more delete the request

completed - [s] all records for this request completed

ready for production - [s] \_allways\_ a human decission!

**RequestMDRecord table**

This table links all the metadata records belonging to a given request.

Repox can only insert records in this table whose request status is “under construction”, if the request is aborted don’t remove links from this table, this task will be done manually.

**MDRecord table**

This table contains the original record and all its refinement. ContentHash will be provided by the Harvester and is used to identify the ESE record. The Repox may only insert new MDRecord and cannot change or delete existing items.

MdRecord states (not a true statemachine idle-procesing can be entered several times)

created - [r] the record is created but not yet processed in any way

idle - nobody is touching it, waiting for more checks

processing - a checker is working on this record

problematic - something went wrong, human intervention might save the record

broken - record is invalid, some check decided this ese is not acceptable

verified - all checks succeeded, could be sent to production

change enrichement state to a boolean enrichment\_check\_done

**URIS Table**

uri state  (a uri record is created by a thread and end as completed or failed)

created - (1) the record is just created not processed

uri verified - (2) the uri responds and returns an OK

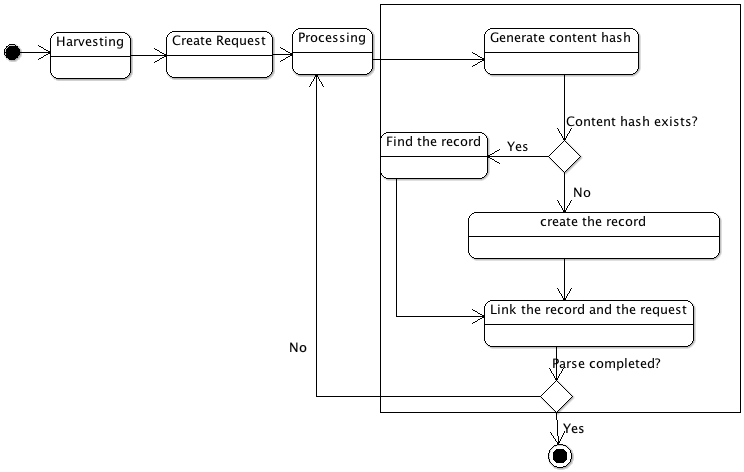
object downloaded - (3)

full\_doc generated - (4)

brief\_doc generated - (5)

completed - (6)

failed - (7) something went wrong, see err\_msg for details in this msg also should be logged what step failed



**Description**

-For each ESE record:

Generate content-hash

- if item identified by content-hash is not found

    - Create new md record, fill it with content-hash and xml dump of ese record

       other fields can be left as defaults

       relevant fields:

         Source\_data - xml dump

 Content\_hash

- Link request and mdrecord identified by content-hash in link table

**Roles**

Remind to describe what stereotypes means here