

Harvesting repository subsets

This use case was submitted as a [Github issue](#) by Kåre Fiedler Christiansen:

At the State and University Library, Denmark, we're harvesting subsets of our repository to different dissemination platforms. Right now, what we are doing is using the RDF triple store to harvest objects with iTQL queries like this:

```
select $object $cm $date
from rmi://localhost/fedora#ri
where
$object info:fedora/fedora-system:def/model#hasModel $cm
and
$cm http://ecm.sourceforge.net/relations/0/2/#isEntryForViewAngle 'SummaVisible'
and
$object http://doms.statsbiblioteket.dk/relations/default/0/1/#isPartOfCollection info:fedora/doms:RadioTV_Collection
and
$object info:fedora/fedora-system:def/model#state info:fedora/fedora-system:def/model#Active
and
$object info:fedora/fedora-system:def/view#lastModifiedDate $date
and
$date http://mulgara.org/mulgara#after '2012-12-12T00:26:56.535Z'^^http://www.w3.org/2001/XMLSchema#dateTime in rmi://localhost/fedora#xsd
order by $date asc
limit 10000
offset 150000
```

(Basically this says "give me active stuff of a specific content model in a specific collection modified since last update, please page it")

The problem is with 2.300.000 object in the repo, this takes hours, and actually fails before returning a result.

Probably the right way to fix this is not to scale the resource index, but to figure a better way to provide this kind of functionality