OAI Sets Generalisation

<?xml version="1.0" encoding="utf-8"?>
<html>
This page is intended to be some initial thoughts on some generalisation work on the concept of OAI sets for DSpace. This work is motivated by the EThOSnet project, which aims to allow deposit of PhD theses into a central hub held at the British Library.

Problem Statement

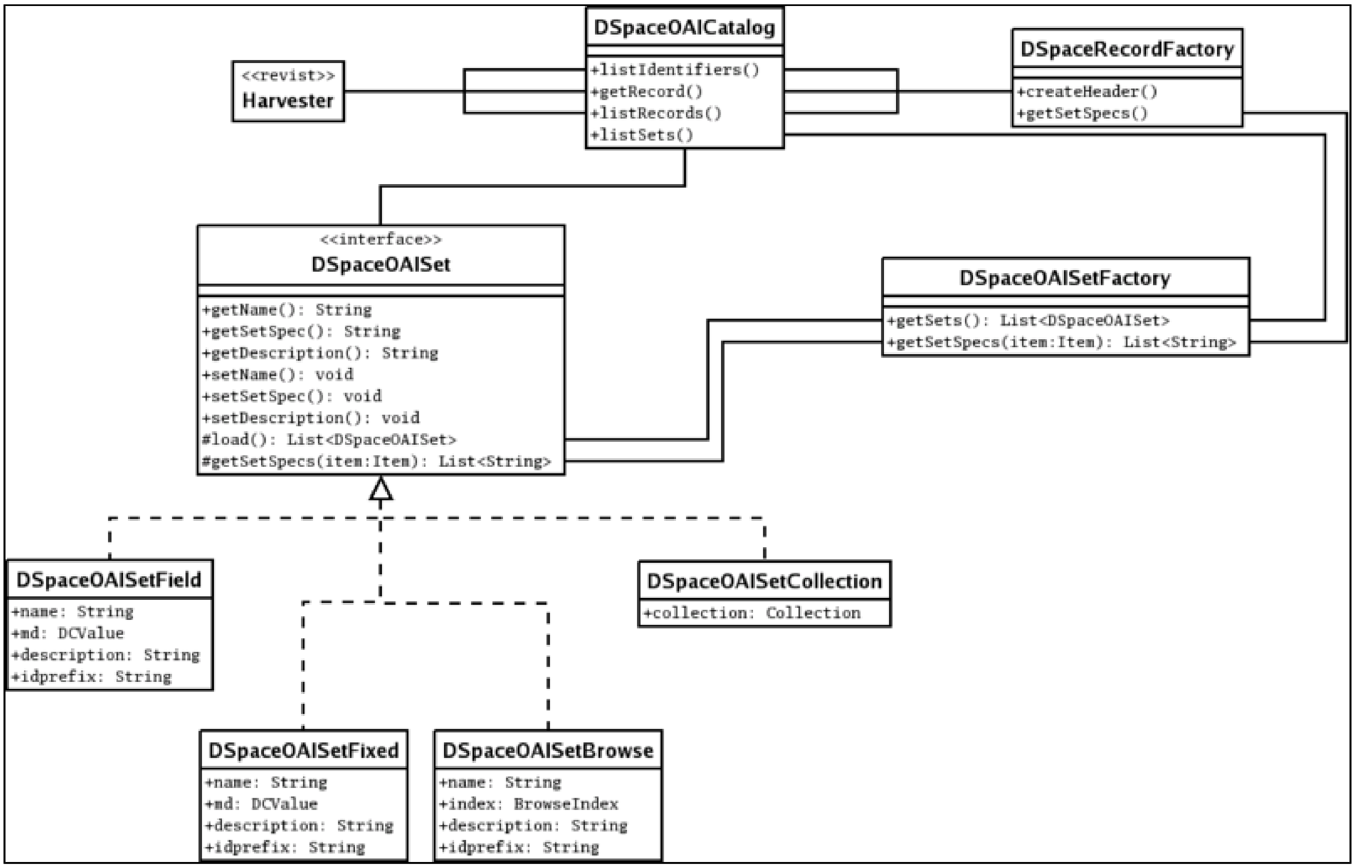
The current issue with DSpace OAI set handling is that it is geared specifically and exclusively to the use of the Collections as the Sets. This creates an artificial relationship between what you expect human users and machine users to want. It also prevents you from having additional sets available to machines which are not available to users.

For the purposes of EThOSnet, it will be necessary to harvest content from repositories (not just DSpace) by content type. This is being done over OAI-PMH, and the requirement is to harvest only theses, filtering them from other content. To harvest from a thesis set is a convenient way of doing this, but this places a particular organisational arrangement on the institution working with EThOS. This is not sufficient for wide adoption, so it is necessary to generalise the process of Set generation and representation within DSpace.

This is a working document looking for a workable solution to the problem.

Design Overview

The following diagram is an outline of the object model proposed for the solution (note that Harvester is not currently thought out).



It introduces a layer of abstraction between the current Set object (a Collection), and the DSpaceOAICatalog. It also allows, then, for the sets to be generated in different ways:

- DSpaceOAISetCollection - this is the current set mechanism, and is a thin wrapper for a Collection
- DSpaceOAISetBrowse - this is a mechanism which will generate sets from distinct browse tables (from the DynamicBrowsePrototype), so if you wanted every single author to be a set in their own right, that would be possible. It would not work for individual titles and so forth.
- DSpaceOAISetFixed - this is a hard coded (in config) set, which returns items with the value of the "name" in the supplied metadata field. For example "Theses or Dissertation" in dc.type
- DSpaceOAISetField - this is effectively the same as DSpaceOAISetBrowse, but doesn't require the DynamicBrowsePrototype. Painfully inefficient, probably won't be implemented.

It is then the job of the DSpaceOAISetFactory to mediate with the list of allowed Set modules. So it will instantiate all the relevant implementations of DSpaceOAISet when requested, and it will also return a list of sets with which an Item is associated when requested. The API for DSpaceOAISet should allow for these operations.

Configuration

Here is a suggested/example set of configurations for the different sets:

```
1. turn on or off collection sets
   oai.set.use_collections = true
```

1. set up by browse index collections

   ```
   a. browse index must be of type "single"
      oai.set.by_browse.<n> = set name:index name:description
   ```

1. set up by - field collections

   ```
   oai.set.by_field.<n> = name:field:description:set spec prefix
   ```

1. set up fixed collections

   ```
   oai.set.fixed.<n> = name:field:description:set spec prefix
   ```

Code Examples

OK, here's some examples as to how this should work in code:

```
// getting a list of all sets
List<DSpaceOAISet> sets = DSpaceOAISetFactory.getSets();
for (DSpaceOAISet set : sets)
{
// generate <setspec>, <setname> and <setdescription> elements
}
```

// looking up an item's set membership

```
DSpaceRecordFactory dsr = getRecordFactory(); // there's some way of doing this
dsr.getSetSpecs(item);
```

</html>