CurationTaskProposal

This idea was implemented in DSpace 1.7 and improved in 1.8+. See Curation System for more details

Issue Addressed

DSpace advertises itself as a preservation-oriented repository, but the default installation contains few tools or services that directly support this claim. Resources do exist, but they are not integrated into the DSpace platform in any straightforward way. Rather, they often live as disconnected code in repositories of research projects and the like.

Proposal

Create a standard way such tools and services could be integrated and used. The idea is to define an abstraction called a 'curation task' which operates at the Item level of the DSpace data model (but whose effects may very well be on individual bitstreams, or in the creation of new ones), and have some generic machinery for managing these tasks (running them, reporting on outcomes, etc). Both the management system and any of the specific curation tasks will be optional, so this curation management system will be an 'add-on', 'module' or whatever we want to call it. It will be designed to be extensible - in that new tasks can be added as they are identified and written. A further important feature of this system will be integration with the workflow system - so that tasks deemed necessary to perform before content is installed into the repository can be accommodated.

Examples of Curation Tasks

While the primary objective is to create a management framework, it only makes sense if there are available tasks to manage. So the initial release should contain several functional, valuable tasks. Some candidate tasks:

- A task to create an Item AIP for replication to a service like DuraSpace
- A format-identification service for uploaded bitstreams (using, e.g. DROID)
- A virus checking service for specific bitstream types
- An obsolete format detector

And there are many others, but these are fairily representative of the sorts of operations under consideration (and for which there is existing work that could be adapted)