

Digital Stacks, How to Store and Manage Your Electronic Copies

Libraries have always been known for their skills in describing, archiving and storing books and journals. Now we see that libraries have to deal with local copies of electronic publications. Describing them the way they have described paper publications can be done in the same way. Storing them on server disks has very often been done by IT departments in various ways. At Wageningen University and Research Library this has resulted in files in numerous directories on server disks. The way it has been organized was reinvented several times again, when new projects lead to numerous new documents.

In 2007 a project was started to organize the storage and to develop a 'digital stacks' application. This application is meant to deal with several problems:

1. The logistic procedures that have to happen, when an electronic copy is offered to or bought by the library.
2. The provision of a persistent URL to students and staff that upload a copy to the libraries digital stacks.
3. The discovery of duplicates of copies.
4. Versioning of electronic copies.
5. Access rights to electronic copies.
6. The storage of electronic copies.

This presentation will explain the problems we had and the solution we developed.

The problem

Almost 8% of all library catalog records describe electronic documents. Close to 10% of all descriptions in our bibliography are about electronic documents. About 25% of all Wageningen UR publications we have described are electronic documents. This goes up to over 50% for descriptions of publications that have been published over the last 5 years.

Even though we can see a clear shift from paper to electronic publications, logistics in the library are still largely focused on paper. It is true that logistics concerning serial administration and document delivery has changed over the years, because of the vast amount of electronic serials and the demand for electronic copies of articles. However, the logistics around the way a publication is received, described and stored has hardly changed over the years.

Publications are ordered and received by the acquisitions department, then go to the cataloging department for a formal description, information specialists add keywords and

categories, call numbers are assigned and copies are shelved. Any changes in these logistics over the last years were focused on tweaking with this chain of activities to speed up the total process to be able to provide the copies to users as fast as possible.

I think our library staff deserves an applause for their flexibility, since so many electronic publications have been described without a proper logistic process !!

On the other hand. This flexibility resulted into mild chaos. Files were placed on various directories on a web server and linked by a URL in the record. Sometimes a link could be constructed from metadata already available in the record and no link was entered at all. Wageningen University Dissertations already contained a dissertation number and electronic dissertations had that number in their file name. Some smaller Dutch publishers provided us with files, that had the necessary meta data in their filename, like volume, issue en page numbers. Electronic versions of Wageningen UR publications were quite often stored and described 2 times. By the cataloging department and by people responsible for the Wageningen Repository.

We needed change !

The solution

And it is a complete rethinking we have done.

The logistic processing of a paper copy of a publication ended with the shelving process.

The logistic process of an electronic copy starts with the shelving process.

The first thing we do with an electronic copy is storing it in our digital stacks which we have called our E-depot. When storing it, only very basic metadata is entered with the upload.

The most important metadata is about the rights to access the document. The other meta data differs for different sort of documents and is added to provide the cataloger with information that can not be found in the actual document. For example educational level (for student theses) or project codes (for institutional publications)

Compound documents like for example a paper plus a power point presentation can be stored interlinked to keep track of the links between these objects.

Once an object is uploaded, the uploader will receive a simple persistent URL for this object: <http://edepot.wur.nl/xxx> where xxx is just a simple number. This URL will always present the user with the stored object, or (in case of a compound document) with an automatically generated 'jump off page'.

Catalogers check new uploads in the depot daily and decide whether they have to be described as a catalog record, institutional repository object or as an article in our documentation database. They make a decent description if not already available and add the identifier for the description to the E-depot record. Now the E-depot records are automatically linked to the descriptions and splash pages for compound documents are given their final headings.

The API

To upload E-depot items, there is a simple web-interface registered suppliers can use but we have also described an API available for applications outside the library environment. The first application that is using this API is the University Student Information System.

Since the university is visited by curriculum evaluation committees it is necessary for them to have easy access to student output. The library is interested in archiving this output, so we decided to offer an API to easily store and retrieve student theses from the Student Information System. When a Student hands over thesis to complete a curriculum not only the grade will be entered in the Student Information System, but also the document itself can be uploaded. Effectively the system uses the API to store the document in the E-depot and stores the persistent URL locally.

This API can be used to integrate other applications or to build for example interfaces to upload using email or applications of small Dutch publishers that use our E-depot to archive their publications.

Integrating traditional serials registration

Since we make arrangements with some publishers to upload electronic versions of their journal articles or report series to our E-depot, we can now see this as subscriptions that are handled by our serials registration system. We will add functionality to the serials administration system to easily check the E-depot for new updates of subscribed titles to enable to check incoming items and start claiming procedures for missing or stagnating items.