

CASESTUDY KWR

AUTOMATIC METADATA IS AN ESSENTIAL OPERATION TO BRIDGE THE KNOWLEDGE GAP



PROBLEM

At the planned digitizing project, documents were difficult to find in OpenText Content Server because of the lack of metadata.

SOLUTION

Automatically add metadata to existing content in Open Text Content Server.

RESULT

Faster access to internal library and increasing the number of available publications available in the online library and, thus, more knowledge made available to the society.

'Bridging Science to Practice' is the motto of KWR, the Dutch institute for water research. One of the ways of KWR sought to increase knowledge transfer between science and practice was to publish knowledge (scientific articles, reports, etc.) on its online library. "The amount of content and the high speed in which it's presented, has not always been what it is now," said Jonie Keessen, librarian at KWR. The start of the construction of the new office building in 2012 was used as a psychological moment to digitize documents.

"All the 170 employees would get a maximum of one meter of bookshelf in the new office. From that moment, we started scanning and digitally storing documents that needed to be preserved and of which we only had a hard copy. They were stored on our central ECM system OpenText Content Server; in other words, the internal digital library."

■ Incomplete metadata

After all documents were scanned and added to the internal digital library, it was time to further optimize the OpenText environment. Keessen explains: "Our wish was to create a connection between the internal digital library and our online library. Public reports and articles that are stored in the internal digital library should be automatically placed on our online library. Previously, that process was manual.

'After automatically adding the metadata our users can find the necessary documents much faster.'

To make the documents searchable, they must have metadata." In practice it turned out that not all content in OpenText had sufficient or the appropriate metadata, so they were not easily accessible for users. For example, it was impossible to search by author or combinations of several words. For this project, KWR asked the help of KBenP combined with Xillio's platform.

■ Automatic metadata

Newly scanned files, as well as very old files from the '60s and '70s, were included in the project. These files were examined using automatic metadata and subsequently metadata was assigned, such as author, title, project number, expertise, client, document type and keywords. "The difficulty was that the older files were often illegible scans and that the layout of the different types of documents has changed over time," said Keessen. "This made it difficult to recognize the document type. Some 200 reports were not OCR'd yet and were added to and OCR'd during the project."

■ Efficient search process

After just three months, the project was completed in June 2016. Longer than initially planned, but that had to do with the diversity in the format of the publications over the years, and many types of documents, and 200 reports that were added to the project.

Keessen talks about the result: "After automatically adding the metadata our users can find the necessary documents much faster. The search process is much more effective. We also added a whole range of older documents from the internal digital library to the online library and, thus, we offer more knowledge. Finally, the underlying process is greatly improved. I used to manually send publications to Communications, which added them to the online library. Now, this is done automatically thanks to the connection between the internal digital library and online library."

■ Continuous improvement

To ensure that new publications are easy to find, project managers add as much metadata as possible when they start a new project in the financial system (AllSolutions). Subsequently, a project folder is automatically created in OpenText CS. All documents and draft publications that are stored in this folder will automatically receive the metadata of the project.

Researchers still send their final publications, such as scientific papers, to Keessen by e-mail. She then adds them to the internal digital library. At the same time, the public publications are automatically published into the online library. This step is being automated further through workflow, so that it's no longer needed to mail publications. This way KWR efficiently shares knowledge with the world despite a small overhead.

Xillio has developed the Content ETL Platform with to solve the challenges organizations face with their fragmented content repositories. The Content ETL Platform consists of a easy scripting language, Xill. It has out-of-the-box connectors to most content repositories, a Unified Data Model and unique modules, like classification and de-duplication, to improve the quality of the metadata. The platform gives organizations the necessary insights in their information landscape, the tools to improve it and the possibility to integrate their point solutions while drastically increasing their agility. Xillio is headquartered in Hilversum, Netherlands. For more information, see www.xillio.com

