

---

# Getting the most out of COCOON: A XML-based webservice for a registration agency

Jan Brase

## Abstract

Since 2005 the German National Library of Science and Technology (TIB) is established as a [DOI](http://www.doi.org) [http://www.doi.org] registration agency for scientific content. Data providers transmit XML-files containing the DC-based metadata descriptions of the scientific data to a webservice infrastructure at the TIB, which was created by the Research center L3S during a project founded by the registration agency for scientific content. Data providers transmit XML-files containing the DC-based metadata descriptions of the scientific data to a webservice infrastructure at the TIB, which was created by the Research center L3S during a project founded by the [German research association \(DFG\)](http://www.dfg.de) [http://www.dfg.de]. This webservice infrastructure is based on the web application framework [COCOON](http://cocoon.apache.org/) [http://cocoon.apache.org/]. We have however extended COCOON with full webservice functionalities. Using XSLT the webservice is furthermore able to transform XML-metadata files into well-formed [PICA](http://oclc.org/pica/) [http://oclc.org/pica/]-files to insert the metadata information into the library catalogue of the TIB.

## Table of Contents

1. Cocoon infrastructure .....	3
2. Cocoon webservice .....	3
3. Status .....	3
4. Conclusion .....	3
Acknowledgements .....	3

# 1. Cocoon infrastructure

Cocoon is an XML publishing framework, it was founded in 1999 as an open source project under Apache Software Foundation. COCOON offers the separation of content, style, logic and management functions in an XML content based web site. This separation allows us to easily change the parts of the architecture or the appearance of the system. Since it is initialised by the retrieval of a XML-file, send to the system by the data providers, every registration starts a XML based pipeline process. This process includes:

- I. Java based registration of the DOI using the [Handle system](http://www.handle.net) [http://www.handle.net].
- II. A XSLT transformation of the XML metadata into well-formed PICA file. PICA is the catalogue format used at the TIB and is based on [MARC21](http://www.loc.gov/marc/) [http://www.loc.gov/marc/], the catalogue system of the Library of congress.
- III. Upload of the PICA file onto a ftp-server to include the metadata into the library catalogue

# 2. Cocoon webservice

The first version of our Cocoon infrastructure used [AXIS](http://ws.apache.org/axis) [http://ws.apache.org/axis] software as a webservice interface. This led to duplication of the code. Since SOAP messages however are written in XML and COCOON is based on XML it was an obvious step to include the webservice functionalities into the system. For the latest version of our infrastructure we have extended COCOON to read and interpret SOAP messages, skipping the AXIS components. First tests have shown a slight gain of time combined with a significant gain of comfort.

# 3. Status

With the infrastructure at the TIB we have registered 250,000 datasets so far (May 2005). Our extended version of COCOON was successfully tested with the last 80,000 datasets.

# 4. Conclusion

We have extended the XML publishing framework COCOON with complete webservice functionalities, using XSLT. We are furthermore able to offer a unique infrastructure for registration tasks, which includes a XSLT formalism to create well formed library catalogue metadata out of XML files containing DC-based metadata.

# Acknowledgements

The extension of COCOON was realised by Jan Hinzmann as a Bachelor Thesis. The registration of data is part of the project "Publication and Citation of Scientific Primary Data", funded by the German Research Foundation (DFG).

# Biography

## Jan Brase

Research coordinator

Research center L3S

Hannover

Niedersachsen

Germany

<http://www.kbs.uni-hannover.de/~brase/>

Jan Brase is coordinating the research issues in the field digital libraries for the [German National Library for Science and Technology \(TIB\)](#) [<http://www.tib-hannover.de>], a member of the L3S. In that context he is technical advising the DOI-registration of scientific content and learning objects at the TIB.