

Title: Requirements to Document and Publication Repositories

Authors: Peter Schirnbacher (Humboldt University)

Abstract: Installing and institutionalizing document and publication repositories at universities creates the opportunity to offer and archive scholarly publications that were produced at the respective universities for a worldwide audience. This new service by the infrastructure institutions library and computing center helps disseminate the idea of electronic publishing as a new tool for academic work.

DINI - the Deutsche Initiative für Netzwerkinformation (German Initiative for Networked Information) - supports this development to reach a higher level of scientific and scholarly communication nationwide and internationally. To reach this goal it is necessary to network document and publication repositories. The publication of electronic theses and dissertations on the internet is a good example for this process. Major prerequisites for success were:

- A nationwide development of metadata standards,
- publication on local repositories through cooperation of libraries and computing centers,
- a workflow for the upload to the German National Library (Die Deutsche Bibliothek), which in turn assumed responsibility for long-term archiving of the theses.

Taking the DINI publication "Electronic Publishing in Higher Education" (March 2002), which was published as a first guideline, as a starting point, this new initiative aims at three targets:

- give a detailed description of the requirements of a certified document and publication repository,
- highlight directions of development in the areas of configuration of repositories and of information exchange through theses repositories,
- offer a certificate visible to both users and operators to certify the adherence to defined standards and recommendations by this repository.

By issuing this certificate, DINI for the first time introduces a quality control for document and publication repositories.

A set of minimum requirements of a repository and its operator mandatory for modern scholarly communication is formulated. At the same time DINI makes

recommendations highlighting foreseeable developments that might turn into future requirements.

A working group within DINI audits the criteria for the DINI certificate against international standards and developments and updates them accordingly.

Title: Project GRACE: A grid based search tool for the global digital library

Authors: Frank Scholze (Stuttgart University), Glenn Haya (Stockholm University), Jens Vigen (CERN) and Petra Prazak (Stuttgart University)

Abstract: The paper will report on the progress of an ongoing EU project called GRACE - Grid Search and Categorization Engine (<http://www.grace-ist.org>). The project participants are CERN, Sheffield Hallam University, Stockholm University, Stuttgart University, GL 2006 and Telecom Italia. The project started in 2002 and will finish in 2005, resulting in a Grid based search engine that will search across a variety of content sources including a number of electronic thesis and dissertation repositories.

The Open Archives Initiative (OAI) is expanding and is clearly an interesting movement for a community advocating open access to ETD. However, the OAI approach alone may not be sufficiently scalable to achieve a truly global ETD Digital Library. Many universities simply offer their collections to the world via their local web services without being part of any federated system for archiving and even those dissertations that are provided with OAI compliant metadata will not necessarily be picked up by a centralized OAI Service Provider as the collection might not be officially registered as an OAI data provider.

GRACE is an attempt to apply an innovative Grid-based solution that will meet the challenges of searching a global heterogeneous collection of documents. The goal of the project is to build a distributed search and categorization engine that will run on the European Data Grid (EDG) and its successor, the Enabling Grids for E-science in Europe (EGEE). The main difference between GRACE and existing search engines is that GRACE has no centralized index. Instead, it will rely on local indexes or search interfaces that are dispersed across web services around the world. These local sources can use different protocols including http, OAI-PMH and Z39.50. In order to include and index even document collections offering no local search possibilities at all, GRACE will use a native search engine based on Lucene. This decentralized approach, along with the scalable processing power provided by the Grid will result in the following advantages to users:

1. Advanced search capabilities which are flexible enough to allow the broadest possible features given the content sources selected for searching.
2. Increased currency of information and indexes.