
Project bwDataDiss: bwData for Dissertations

Tobias Kurze^{*1}, Wiebke Beckmann^{*2}, and Matthias Bonn^{*3}

¹Karlsruhe Institute of Technology [Karlsruhe] (KIT) – P.O. Box 3640 76021 Karlsruhe, Germany

²University of Freiburg [Freiburg] – Platz der Universität 2, 79098 Freiburg, Germany

³Karlsruhe Institute of Technology [Eggenstein-Leopoldshafen] (KIT) – Campus North
Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen Campus South Kaiserstraße 12 76131
Karlsruhe, Germany

Abstract

The goal of bwDataDiss is to build up digital infrastructure for PhD students and university libraries in the state of Baden-Württemberg in Germany to archive and enable access to research data in the context of doctoral dissertations.

bwDataDiss is a three year project funded by the Ministry for Science and Art of Baden-Württemberg. Project partners are both the university libraries and computing centres of Freiburg and Karlsruhe.

In the context of their doctoral dissertation, Ph.D. students often produce research data. As the scientific society becomes more and more aware of the importance of verification of research results, the general need to build up digital infrastructures to archive and enable access to research data arises. But currently libraries often lack the digital infrastructure to handle research data as they are often heterogeneous with regard to contents and filetypes. Furthermore the amount of research data varies strongly within research fields.

bwDataDiss enables university libraries in the State of Baden-Württemberg to archive research data together with the final dissertation. Via the institutional repository the user (=Ph.D. student) is able to upload research data to bwDataDiss. To guarantee integrity, checksums are calculated every time the data is transferred.

Archived research data can be accessed through a web portal by the general public. They are described by a set of metadata, which in turn is gathered by the libraries and kept in sync with bwDataDiss. The bibliographic metadata scheme is of general nature and not community specific.

The project promotes and is committed to Open Access, but also allows to arrange 'embargos' i.e. time spans during which no public access is allowed. As the DFG (german research funding agency) demands, bwDataDiss preserves research data for at least ten years, operates quality checks on the files and makes them accessible to other researchers and the general public. Another main objective of the project is, that bwDataDiss should be as easy as possible to use both for Ph.D. students and the staff of the library.

As bwDataDiss must deal with different library systems it has a certain flexibility regarding the integration with those systems. Besides a web-interface bwDataDiss also provides

*Speaker

an API to allow an almost seamless integration and uses the Baden-Württemberg identity management federation (bwIDM) to provide SAML-based web single-sign-on user authentication. Uploads of files up to a size of 10GiB will be supported. For this, bwDataDiss relies on hierarchical and flexible storage services from the SCC Data centre at KIT and cooperates in the context of the bwDataArchiv project. Before the data is actually stored in the archive, a so-called ‘Characterization’ of the research data is performed, i.e. the type of each file is determined. This supports the assessment of the submitted data and helps with future curation of the research data.

bwDataDiss is planned to go live by the end of the year.

On the author:

Tobias Kurze studied computer sciences at KIT (Uni Karlsruhe) and INSA Lyon. Former research associate at (KIT -) SCC (Steinbuch Centre for Computing) with focus on distributed, grid and cloud computing. Now research associate at the KIT library and in charge of project bwDataDiss (long term storage of research data), which he will present at ETD2016.

Keywords: project, library, dissertations, archive