

Webarchivierung im BSZ

Dr. Barbara Löhle
Bibliotheksservice-Zentrum Baden-Württemberg
Göppingen, 2. Mai 2012

- Webarchivierung im BSZ als Teil der Langzeitarchivierung mittels **SWBcontent**
- **SWBcontent** für Bibliotheken
- **SWBcontent** für Archive
- **SWBregio** – Webseitenarchiv für Kommunen
- Ergebnisse der Webarchivierung mittels **HTTrack** und **Heritrix 3.x** in **SWBregio**
- Rechteverwaltung
- Aktuell: Volltextsuche mittels **Apache Solr**

- **Baden-Württembergisches Online-Archiv (BOA) als erste Web-Applikation auf der Basis von SWBcontent**

sammelt,

erschließt,

präsentiert,

archiviert

[Webseiten](#) und Publikationen aus und über
Baden-Württemberg

- Kooperation von Landesarchiv, Landesbibliotheken und BSZ
 - LA: [Webseiten von Landeseinrichtungen](#)
 - WLB und BLB sammeln im Rahmen Landesbibliographie und Pflichtexemplar.
- Seit 2004 im Betrieb
- Technische Grundlage: Software SWBcontent des BSZ

- SWBcontent liegt in 2 Varianten vor (unterschiedliche beschreibende Metadaten):
 - ⤴ Archive harvesten Homepages, d.h. umfangreiche Downloads
 - ⤴ Bibliotheken harvesten einzelne PDF-Dokumente als Monographien oder als Teile von mehrbändigen Werken (Extremfall: Dokumentenserver auf Basis von PDFs)
- Die Durchführung von Web-Downloads einzelner PDF-Files und ganzer Websites erfolgt durch Integration der Open Source Webarchivierungs-Software, vgl. (Bundeskonzferenz der Kommunalarchive beim Deutschen Städtetag (BKK)
 - ⤴ HTTrack
 - ⤴ Heritrix 3.x
- Der Upload von einzelnen PDF-Files ist ebenso möglich.

Web literatur-im-netz.dla-marbach.de

deutsches literatur archiv marbach

Literatur im Netz

Home Kontakt Hilfe

BSZ

SaarDok - Der Archivserver für das Saarland - Opera

SaarDok - Der Arc...

Web saardok.sulb.uni-saarland.de

SaarDok

Der Archivserver für das Saarland

Saarländische Universitäts- und Landesbibliothek

Home Kontakt Hilfe

BSZ

Informations- und Bibliotheksportal des Bundes - Opera

Informations- und ...

Web www.publikationsserver-bund.de

Das Bibliotheksportal

Informations- und Bibliotheksportal des Bundes

EINFACHE SUCHE
ERWEITERTE SUCHE
BROWSEN
EDITIEREN

Recherche:

Suche starten

Web www.boa-bw.de/menu.do

BADISCHE LANDESBIBLIOTHEK
LANDESARCHIV BADEN-WÜRTTEMBERG
WÜRTTEMBERGISCHE LANDESBIBLIOTHEK

BOA BOV Baden-Württembergisches Online-Archiv Landesbibliotheken

Home Kontakt Hilfe

BSZ

SWBdok - Opera

SWBdok

Web swbdok.bsz-bw.de

SWBdok

Digitale Publikationen Online

Home Kontakt Hilfe

BSZ

EINFACHE SUCHE
ERWEITERTE SUCHE
BROWSEN
EDITIEREN

Recherche: Suche starten

SWBdok

In SWBdok werden freie Internetpublikationen gesammelt, erschlossen, der Öffentlichkeit zur Verfügung gestellt und dauerhaft archiviert.

Teilnehmerbibliotheken des Südwestdeutschen Bibliotheksverbundes (SWB) können z.B. E-Books, die für Forschung und Lehre notwendig sind, hier einstellen und derzeit frei zur Verfügung stellen. Die Publikationen sind in der Datenbank des SWB nachgewiesen und auch in den Katalogen der einstellenden Bibliotheken enthalten. Über eine URL erfolgt die Verbindung zur archivierten Publikation.

SWBdok kann von allen Teilnehmerbibliotheken des Südwestdeutschen Bibliotheksverbundes genutzt werden. Voraussetzung ist, dass die Rechte der Publizierenden gewahrt werden. Dazu verpflichten sich die Bibliotheken, die Inhalte in SWBdok einstellen.

Als Pilotpartner stellt die Bibliothek der HTWG Konstanz Publikationen aus dem Internet in

- Extremfall eines reinen Dokumentenservers
 - ⤴ SWBdok
 - ⤴ Publikationsserver im Informations- und Bibliotheksportal des Bundes (nicht allgemein zugänglich)
- Extremfall einer „Archiv-Sammlung“ bestehend einzig aus Downloads von Web-Seiten als Netzliteratur, Weblogs und Zeitschriften: „Literatur im Netz“ des Literaturarchivs Marbach
- heterogene Sammlungen (PDF-Dokumente und Web-Seiten) im Fall von
 - ⤴ BOA Landesbibliotheken
 - ⤴ SaarDok, der Archivserver für das Saarland

The screenshot displays the SWBregio website in an Opera browser window. The browser's address bar shows the URL 'www.swbregio.de'. The page layout includes a search bar with a 'Suche starten' button and a navigation menu with options like 'EINFACHE SUCHE', 'ERWEITERTE SUCHE', 'BROWSEN', and 'EDITIEREN'. The page content includes the title 'SWBregio Digitales Webseitenarchiv' and a description of the digital archive project.

- Als Kooperation von Landesarchiv Baden-Württemberg und BSZ stellt SWBregio eine Plattform dar zur Archivierung von Webseiten im regionalen Bereich, d.h. von
 - Städten
 - Kommunen
 - Kreisen
 - und den damit verbundenen Körperschaften
- Das Angebot ist nicht auf Baden-Württemberg beschränkt, z.B. Stadt Speyer.
- BKK verweist in ihrer „Empfehlung zur Speicherung kommunaler Webseiten – Teil 2: Technik“ auf das Serviceangebot des BSZ als „empfehlenswertes Beispiel“ eines komplett Web-gestützten Workflows zur Archivierung der Webseiten von Kommunalarchiven

The screenshot shows a web browser window displaying the SWBregio website. The browser's address bar shows the URL www.swbregio.de/menu.do;jsessionid=EBB66BEC7A9222F2ADF35A9. The website has a green header with the SWBregio logo and navigation links: Home, Kontakt, and Hilfe. Below the header is a green banner with the text "Digitales Webseitenarchiv". The main content area is white and contains a search section with links for "EINFACHE SUCHE", "ERWEITERTE SUCHE", "BROWSEN", and "EDITIEREN". Below this is a section titled "Browsen über Dokumentarten" with a list of categories and their object counts:

- 01 Homepage ([45 Objekte](#))
- 02 thematische Website ([3 Objekte](#))
- 03 BSZ ([1 Objekt](#))
- 04 Stadt Mannheim ([42 Objekte](#))
- 05 Stadt Speyer
- 06 Stadt Tübingen ([1 Objekt](#))
- 07 Stadt Villingen-Schwenningen ([1 Objekt](#))
- 08 Sonstiges

SWBregio

Home
Kontakt
Hilfe



[Titelsatz](#) >> [Präsentation](#) >> [Vollbildanzeige](#)

Die Anzeige dieser Ressource stammt aus "SWBregio".
Sie wurde am **11.05.2011** archiviert.

ISG STADTARCHIV MANNHEIM
INSTITUT FÜR STADTGESCHICHTE

[HOME]

- AKTUELLES
- ERSTINFORMATIONEN
- DIENSTLEISTUNGEN
- ONLINE-RECHERCHEN
- VERÖFFENTLICHUNGEN
- FAQS ZUR STADTGESCHICHTE
- STADTPUNKTE - MANNHEIMER GESCHICHTE VOR ORT
- RÜCKBLICKE
- KZ GEDENKSTÄTTE SANDHOFEN
- MANNHEIMER ARCHITEKTUR- UND BAUARCHIV E.V. (MAB)
- VEREIN DER FREUNDE DES STADTARCHIVS E.V. (VFS)
- VERKAUFSSHOP
- ARCHIV FÜR KIDS
- IMPRESSUM

Unsere aktuellen Neuerscheinungen

Landerer, Helmut:
"daß die Wittib oft besoffen sei!" Eine Quellenstudie zu den Mannheimer Küfern anhand der Verlassenschaftsakten des 17. bis 19. Jahrhunderts. Mannheim 2011. 115 S.
ISBN 978-3-9813584-2-1

Sterben und Erben war schon immer ein Thema – ging es ums liebe Geld, dann stritten sich die Hinterbliebenen, und die Behörden sicherten sich die Erbschaftssteuer. Oder sie griffen ein, wenn dem Vermögen der Waisenkinder Gefahr drohte, weil es sprichwörtlich „versoffen“ wurde, wie der tragische Fall einer Witwe zeigt. [Mehr](#)

14,80 € (9 € für Mitglieder des Vereins der Freunde des Stadtarchivs)

[Zum Bestellformular](#)

Seidel, Winfried A.:
„Tatort“ Mannheim. Wie das Automobil entstand. Mannheim 2011. 120 S.
ISBN 978-3-9813584-1-4

„Tatort“ Mannheim - wie das Automobil entstand erzählt über die aus Baden und aus Württemberg stammenden Pioniere Carl Benz, Gottlieb Daimler und Wilhelm Maybach. Sie gingen sehr unterschiedliche Wege, dennoch hatten sie alle ein Ziel: Die Menschen mobil zu machen. [Mehr](#)

29,80 € (20 € für Mitglieder des Vereins der Freunde des Stadtarchivs)



SWBregio

Home
Kontakt
Hilfe



[Titelsatz](#) >> [Präsentation](#) >> [Vollbildanzeige](#)

Die Anzeige dieser Ressource stammt aus "SWBregio".
Sie wurde am **20.07.2011** archiviert.

ISG STADTARCHIV MANNHEIM
INSTITUT FÜR STADTGESCHICHTE

[HOME]

- AKTUELLES
- ERSTINFORMATIONEN
- DIENSTLEISTUNGEN
- ONLINE-RECHERCHEN
- VERÖFFENTLICHUNGEN
- FAQS ZUR STADTGESCHICHTE
- STADTPUNKTE - MANNHEIMER
GESCHICHTE VOR ORT
- RÜCKBLICKE
- KZ GEDENKSTÄTTE SANDHOFEN
- MANNHEIMER ARCHITEKTUR- UND
BAUARCHIV E.V. (MAB)
- VEREIN DER FREUNDE DES
STADTARCHIVS E.V. (VFS)
- VERKAUFSSHOP
- ARCHIV FÜR KIDS
- IMPRESSUM

Liebe Besucher und Freunde des Stadtarchivs,

unser neues
Veranstaltungsprogramm
erscheint im Frühherbst 2011.

Wir hoffen, dass wir mit unserer
Themenauswahl und unseren
geplanten Veröffentlichungen
erneut Ihr Interesse werden
wecken können und freuen uns
auf Ihren Besuch!

Schon jetzt möchten wir Sie
aufmerksam machen auf unser

**Archivfest am 17.
September 2011 von
11 bis 16 Uhr**

Ganz im Zeichen von
Aufbruch und Neuanfang
steht es in diesem Jahr unter
dem Motto
"Rad und Tat".

Mit einer Theaterpremiere
lassen wir zwei Pioniere
wiederauferstehen, die voller
Tatendrang und mit viel
Erfindungsgeist versuchten, die

weitere Infos:
www.stadtarchiv-mannheim.de

RAD UND TAT
CARL BENZ & FRIEDRICH HECKER

ARCHIVFEST 2011
Theater | Musik | Führungen | Information

Samstag, 17. September 2011, 11-16 Uhr
Stadtarchiv Mannheim-ISG, Collini-Center

STADTARCHIV MANNHEIM
 INSTITUT FÜR STADTGESCHICHTE

INTERNET ARCHIVE

Go

JAN FEB MRZ
14
 2011 2012 2013

[Close](#) ✕
[Help](#) ?

[HOME]

AKTUELLES

ERSTINFORMATIONEN

DIENSTLEISTUNGEN

ONLINE-RECHERCHEN

VERÖFFENTLICHUNGEN

FAQS ZUR STADTGESCHICHTE

STADTPUNKTE - MANNHEIMER
GESCHICHTE VOR ORT

RÜCKBLICKE

KZ GEDENKSTÄTTE SANDHOFEN

MANNHEIMER ARCHITEKTUR- UND
BAUARCHIV E.V. (MAB)

VEREIN DER FREUNDE DES
STADTARCHIVS E.V. (VFS)

VERKAUFSSHOP

ARCHIV FÜR KIDS

IMPRESSUM

Tag der Archive, 3.3. 2012

Vortrag
Mannheim – damals – gestern – heute
Familienforschung einmal anders
 Von Karen Strobel und Nicole Kane
 14 Uhr, Friedrich-Walter-Saal, Collini-Center, Erdgeschoss

Die Geschichte einer Stadt ist auch die Geschichte ihrer Häuser, Menschen und Familien, die darin leben. Mit diesem Vortrag wollen wir Sie ein wenig in die Vergangenheit entführen... [Mehr](#)

„MITTWOCHS BEIM ARCHIV“

Vortrag
Dr. Dirk Hecht (Stadtarchiv Schriesheim)

Von der Friedrichsburg in Mannheim ins Rathaus von Schriesheim.
Die Geschichte der wohl ältesten erhaltenen Turmuhr Deutschlands.

Mittwoch 15. Februar 2012, 19 Uhr
 Friedrich-Walter-Saal des Stadtarchivs-ISG,
 Collini-Center, Erdgeschoss

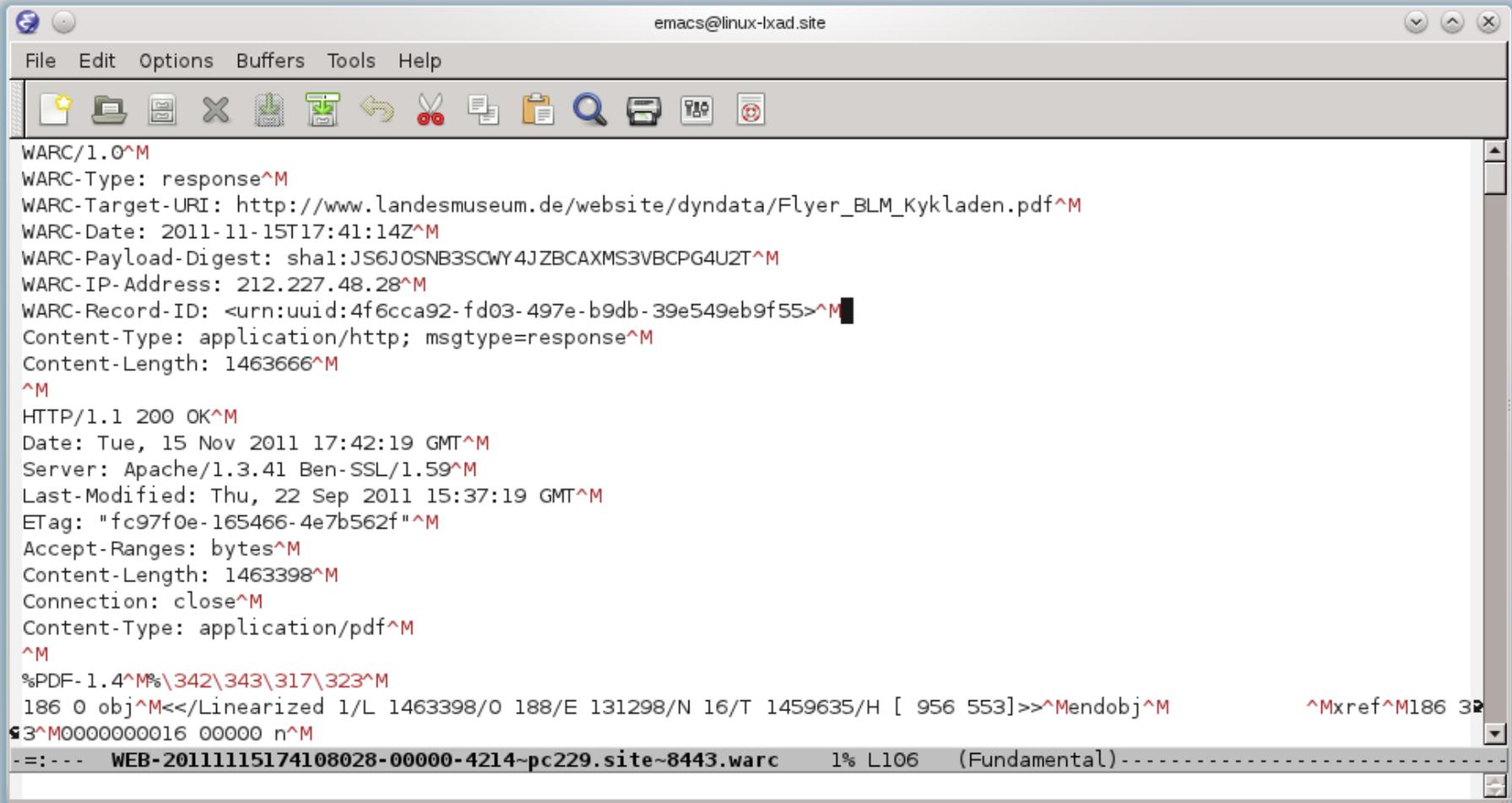
Nach den großen Zerstörungen 1674 durch die Franzosen brauchte Schriesheim nicht nur ein neues Rathaus, sondern auch eine neue Rathausuhr. So kaufte die Gemeinde Schriesheim

[MANNHEIM]

STADT MANNHEIM
Leben im Quadrat.

- Komponenten im Fall von Heritrix 3.x
 - ⤴ Heritrix 3.x Crawl-Server (erzeugt Downloads im WARC-Format)
 - ⤴ Darstellung mittels der Wayback-Machine (eine weitere Open Source Komponente, deren Design noch angepaßt werden muß.)
- zum WARC (Web ARChive) Format:
 - ⤴ stellt seit 2009 einen ISO-Standard dar: ISO 28500:9000
 - ⤴ Kein xml-Format, sondern ein “binary-embedded” strukturierter Text
 - ⤴ Aufzeichnung der gesamten Netzwerk-Kommunkation bei Durchführung einesWeb- Downloads
 - ⤴ Für die Zukunftssicherheit der Web-Downloads von SWBcontent ist es wichtig WARC einzusetzen.

Zentraler WARC Datensatz mit Heritrix 3.1:



```
emacs@linux-lxad.site
File Edit Options Buffers Tools Help
WARC/1.0^M
WARC-Type: response^M
WARC-Target-URI: http://www.landesmuseum.de/website/dyndata/Flyer_BLM_Kykladen.pdf^M
WARC-Date: 2011-11-15T17:41:14Z^M
WARC-Payload-Digest: sha1:JS6JOSNB3SCWY4JZBCAXMS3VBPCG4U2T^M
WARC-IP-Address: 212.227.48.28^M
WARC-Record-ID: <urn:uuid:4f6cca92-fd03-497e-b9db-39e549eb9f55>^M
Content-Type: application/http; msgtype=response^M
Content-Length: 1463666^M
^M
HTTP/1.1 200 OK^M
Date: Tue, 15 Nov 2011 17:42:19 GMT^M
Server: Apache/1.3.41 Ben-SSL/1.59^M
Last-Modified: Thu, 22 Sep 2011 15:37:19 GMT^M
ETag: "fc97f0e-165466-4e7b562f"^M
Accept-Ranges: bytes^M
Content-Length: 1463398^M
Connection: close^M
Content-Type: application/pdf^M
^M
%PDF-1.4^M%\342\343\317\323^M
186 0 obj^M<</Linearized 1/L 1463398/0 188/E 131298/N 16/T 1459635/H [ 956 553]>>^Mendobj^M      ^Mxref^M186 32
3^M0000000016 00000 n^M
--:--- WEB-20111115174108028-00000-4214~pc229.site~8443.warc 1% L106 (Fundamental)-----
```



- Vergabe der Zugriffsrechte auf die Objekte (hier: der Downloads)
 - Free - frei zugängliche Objekte
 - Blocked - nur im Login-Modus zugängliche Objekte
 - Moving Wall - Blocked Objekte, die nach einem bestimmten Zeitpunkt in den Status Free übergehen
 - Domain_all - Beschränkung des Zugriffs auf Objekte auf Clients bestimmter IP-Ranges.

- Volltextsuche mittels Apache Solr:
 - Bisherige Suche in SWBcontent: Metadatenuche
 - nun Einsatz einer weiteren Server Komponente: Apache Solr (Searching on Lucene w/Replication)
 - Integration für den Upload von einzelnen PDF-Files
 - in Kürze: Erweiterung auf den Fall allgemeiner Web-Downloads, d.h. Files mit jenen File-Formaten werden indexiert, die mittels Solr-Bordmitteln identifiziert und extrahiert werden können.
 - Bsp.: SWBdok

Home
Kontakt
Hilfe

BSZ |

SWBdok
Digitale Publikationen Online

EINFACHE SUCHE
ERWEITERTE SUCHE
BROWSEN
EDITIEREN

Metadaten-Recherche

Titel

Personen und Körperschaften

Notationen und Schlagwörter

Verknüpfung zwischen den Begriffen: and

Volltextsuche

Recherche:

```
-<result name="response" numFound="1" start="0">
  -<doc>
    -<arr name="content_type">
      <str>application/pdf</str>
    </arr>
    -<str name="id">
      http://swbcontest.bsz-bw.de/swbdok/downloads
      /frei/b136e7e5-9f61-4b45-9ed8-6676ee2f3c6e/0/campusplan_htwg_01.pdf
    </str>
    <date name="last_modified">2009-11-12T15:27:34Z</date>
    <int name="swb_object_id">11</int>
    -<str name="swb_object_link">
      http://swbcontest.bsz-bw.de/swbdok
      /b136e7e5-9f61-4b45-9ed8-6676ee2f3c6e.html
    </str>
    <int name="swb_root_object_id">10</int>
    -<str name="swb_root_object_link">
      http://swbcontest.bsz-bw.de/swbdok/dnb946728372.html
    </str>
    -<arr name="text">
      -<str>
        Campusplan Hochschule Konstanz - Technik, Wirtschaft und Gestaltung |
        Brauneggerstr. 55 | D-78462 Konstanz Tel.: + 49 7531 206-0 | Fax: + 49 7531
        206-400 | kontakt@htwg-konstanz.de | www.htwg-konstanz.de
      </str>
    </arr>
  </doc>
</result>
</response>
```

The screenshot shows the SWBdok search interface. At the top left, the logo 'SWBdok' is displayed above the text 'Digitale Publikationen Online'. To the right, there are navigation links: 'Home', 'Kontakt', and 'Hilfe', along with the 'BSZ' logo. Below the navigation, there are four search options: 'EINFACHE SUCHE', 'ERWEITERTE SUCHE', 'BROWSEN', and 'EDITIEREN'. The main content area is titled 'Ergebnis der Volltextsuche' and shows the search criteria 'für den Suchbegriff "'Campusplan Hochschule Konstanz'" - Treffer: 1'. A search result is displayed in a box with a yellow border, containing the title 'Wie ein Hirte seine Herde hütet ... ; 1' and the number of hits '- Treffer: 1'. Below the title, there are two links: '>> zur Titelaufnahme' and '>> zum archivierten Objekt'. The result itself is 'Campusplan Hochschule Konstanz - Technik, Wirtschaft und ...'.

Home
Kontakt
Hilfe

BSZ |

SWBdok
Digitale Publikationen Online

EINFACHE SUCHE
ERWEITERTE SUCHE
BROWSEN
EDITIEREN

Ergebnis der Volltextsuche

für den Suchbegriff "'Campusplan Hochschule Konstanz'" - Treffer: 1

▼ **Wie ein Hirte seine Herde hütet ... ; 1** - Treffer: 1

>> zur Titelaufnahme
>> zum archivierten Objekt

● **Campusplan Hochschule Konstanz** - Technik, Wirtschaft und

Campusplan

HOCHSCHULE
KONSTANZ
TECHNIK, WIRTSCHAFT
UND GESTALTUNG

Seerhein

Villa Rheinburg

Sportplatz

P

H I K

Forum

A M

G B

P F C

E D

Wiesental
Lorenzstraße
Hochschulstraße
Seerheinstraße

Hochschule Konstanz - Technik, Wirtschaft und Gestaltung | Brauneggerstr. 55 | D-78462 Konstanz
Tel.: + 49 7531 206-0 | Fax: + 49 7531 206-400 | kontakt@htwg-konstanz.de | www.htwg-konstanz.de

Zum Abschluß:

Danke für die Aufmerksamkeit

Fragen und Kommentare?