BSZ Bibliotheksservice-Zentrum Baden-Württemberg

Webharvesting als Dienstleistung

Renate Hannemann | nestor for newbies | 15.05.2018

BSZ:

- seit 2004 Dienstleistung für Webharvesting mit Eigenentwicklung SWBcontent:
  - 3 LBs: Stuttgart, Karlsruhe, Saarbrücken
  - Archive: Landesarchiv BW, Dt. Literaturarchiv, Kommunalarchive
  - 6 Installationen mit Webharvesting-Komponente
- rd. 1.700 Webpräsenzen mit rd. 12.000 Snapshots
- Größtmögliche Vollständigkeit, hohe Spiegelungstiefe, -genauigkeit und Authentizität
- selektives Harvesting insbes. institutioneller Websites und Event-Harvesting

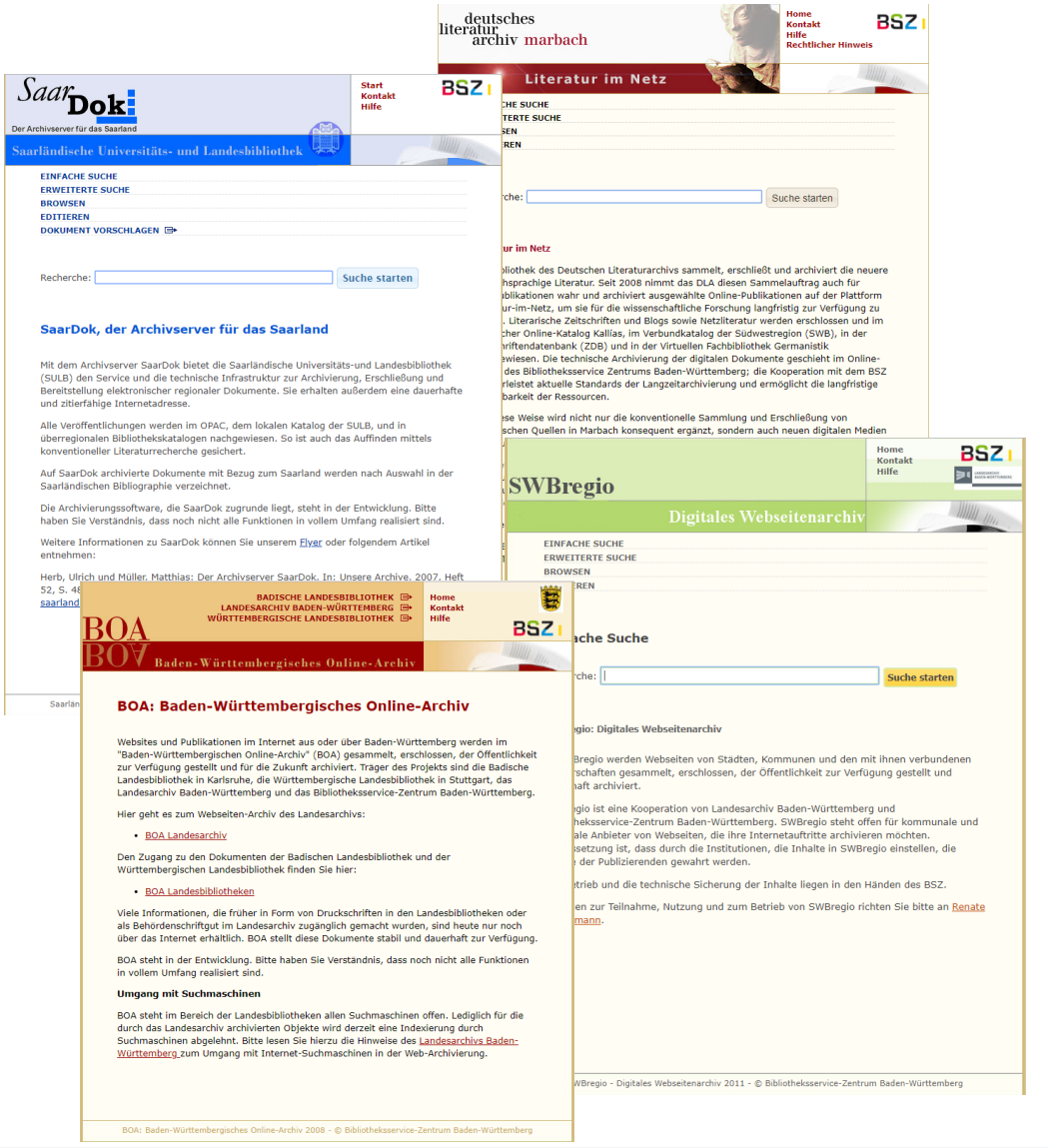
BSZ Bibliotheksservice-Zentrum Baden-Württemberg

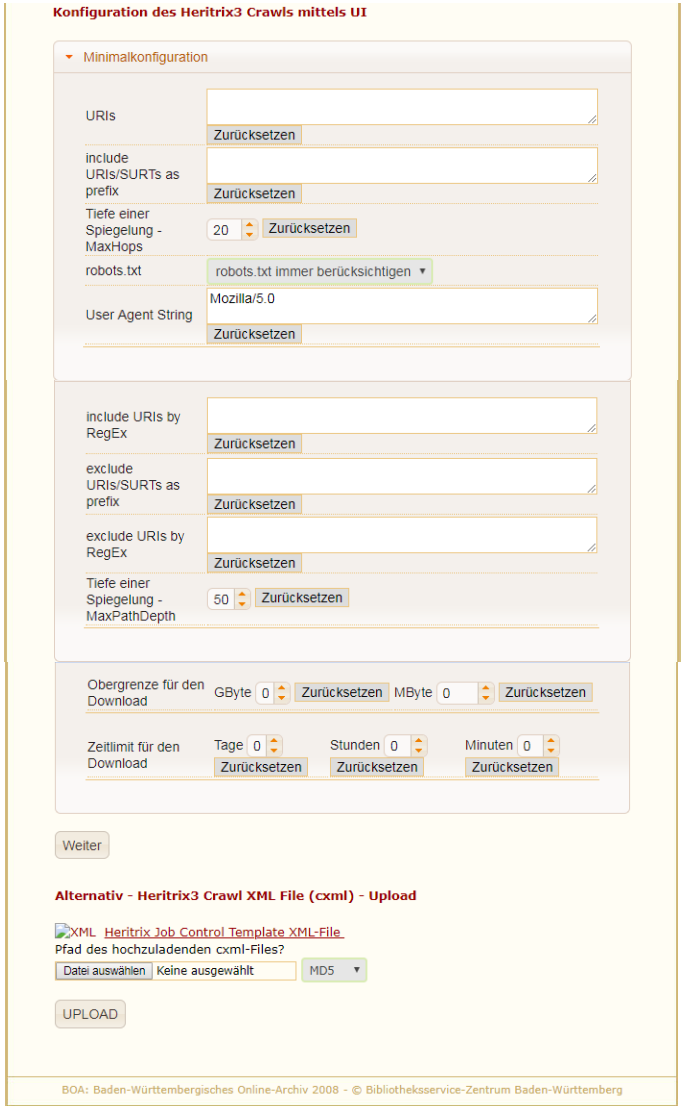
SWBcontent (2004 – 2017/2018) :

- Oberfläche in deutscher Sprache
- bibliograph. Metadaten in SWBcontent oder Übernahme aus SWB
- Parametrisierung, Crawling und Qualitätskontrolle (QA)
- HTTrack, ab 2012 Heritrix3
- Recherche, Präsentation in SWBcontent / Viewer
- zuverlässiges Tool mit überschaubarem Komfort bei Reporting, Logging
- keine automatischen Crawls (Scheduling)
- Erschließung und Crawling autonom durch Kunden

Webharvesting als Dienstleistung dezentrales Crawling

Bibliotheksservice-Zentrum Baden-Württemberg





BSZ Bibliotheksservice-Zentrum Baden-Württemberg

Aufwände BSZ und Erfahrungen:

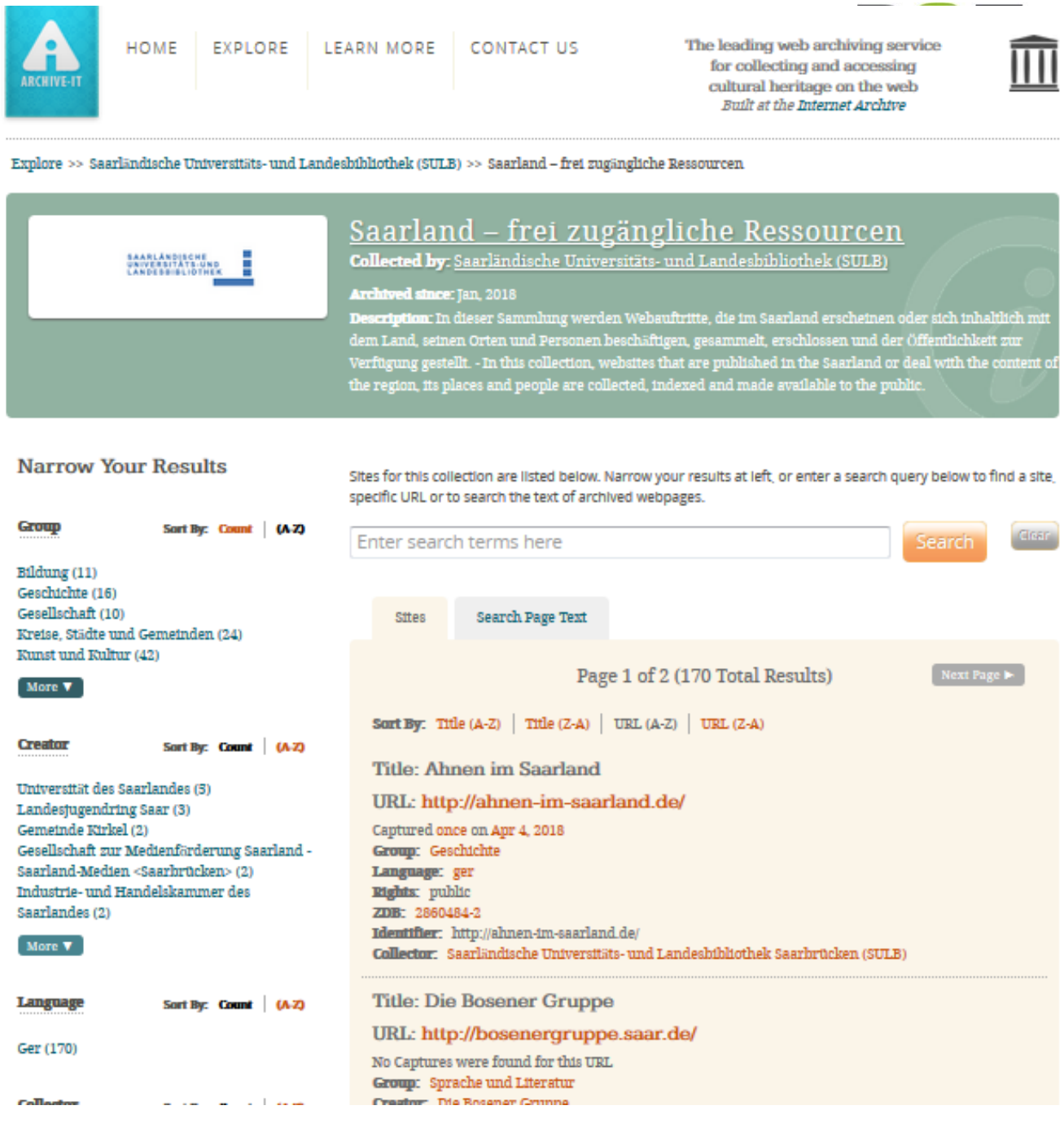
- Administration und Entwicklung Applikation, Viewer
- Administration Präsentationsdaten im BSZ
- Technische Infrastruktur, Speicherinfrastruktur lokal im BSZ und Landesspeicher (LSDF am KIT Karlsruhe)
- Support: Schulungen, Dokumentation, Parametrisierung Spezialfälle, Problem- und Fehleranalyse
- hohes Kommunikationsaufkommen mit Kunden, zeitintensive Fehlerbehandlung
- Datenqualität: schwankend und nicht beeinflussbar

Entwicklung im BSZ seit 2016:

- Ablösung SWBcontent
- Entwicklung eigenes Tool?
- Nutzung Fremdsystem? Evaluation
- Nutzung Dienstleister Archive-It, vollständige Migration der Altdaten
- neues Servicemodell

Archive-It

- seit 2006 Angebot des Internet Archive in S.F.
- rd. 500 Partner weltweit
- komfortables Workflow-Tool:
  Erschließung, Parametrisierung, Jobmanagement, Zugriffsschutz, QA, Schnittstellen
- Heritrix3 + Eigenentwicklungen (Umbra, Brozzler)
- Individualisierbare Präsentation der archivierten Ressourcen als eigenständige Archive im Web
- Volltextrecherche (Websites und pdf)
- Oberfläche, Dokumentation, Support in englischer Sprache



ARCHIVE-IT

HOME | EXPLORE | LEARN MORE | CONTACT US

The leading web archiving service for collecting and accessing cultural heritage on the web
Built at the Internet Archive

Explore >> Saarländische Universitäts- und Landesbibliothek (SULB) >> Saarland – frei zugängliche Ressourcen

Saarland – frei zugängliche Ressourcen

Collected by: Saarländische Universitäts- und Landesbibliothek (SULB)

Archived since: Jan, 2018

Description: In dieser Sammlung werden Webauftritte, die im Saarland erscheinen oder sich inhaltlich mit dem Land, seinen Orten und Personen beschäftigen, gesammelt, erschlossen und der Öffentlichkeit zur Verfügung gestellt. - In this collection, websites that are published in the Saarland or deal with the content of the region, its places and people are collected, indexed and made available to the public.

Narrow Your Results

Group Sort By: Count | (A-Z)

Bildung (11)
Geschichte (16)
Gesellschaft (10)
Kreise, Städte und Gemeinden (24)
Kunst und Kultur (42)

More ▼

Creator Sort By: Count | (A-Z)

Universität des Saarlandes (5)
Landesjugendring Saar (3)
Gemeinde Kirkel (2)
Gesellschaft zur Medienförderung Saarland - Saarland-Medien <Saarbrücken> (2)
Industrie- und Handelskammer des Saarlandes (2)

More ▼

Language Sort By: Count | (A-Z)

Ger (170)

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Enter search terms here | Search | Clear

Sites | Search Page Text

Page 1 of 2 (170 Total Results) Next Page ►

Sort By: Title (A-Z) | Title (Z-A) | URL (A-Z) | URL (Z-A)

Title: Ahnen im Saarland

URL: http://ahnen-im-saarland.de/

Captured once on Apr 4, 2018
Group: Geschichte
Language: ger
Rights: public
ZDB: 2860484-2
Identifier: http://ahnen-im-saarland.de/
Collector: Saarländische Universitäts- und Landesbibliothek Saarbrücken (SULB)

Title: Die Bosener Gruppe

URL: http://bosenergruppe.saar.de/

No Captures were found for this URL
Group: Sprache und Literatur
Creator: Die Bosener Gruppe

BSZ Bibliotheksservice-Zentrum Baden-Württemberg

„SWBregio“ seit 2017

- Zentraler Crawlingservice durch BSZ
- Konsortiallösung für derz. 20 Kommunal- und Kreisarchive
- Auftragscrawling nach schriftlicher Beauftragung
- rd. 320 Websites
- elastisches Preismodell nach Nutzungsintensität
- Archiv: Spiegelungserlaubnis; Beauftragung zu Crawling und Präsentation der Ressource
- BSZ: Einrichten; Metadaten in Archive-It; Parametrisierung; Tests; produktiver Betrieb und QA
- Recherche, Präsentation in Archive-It
- Archiv hat keinen Zugriff auf das Workflow-Tool

BSZ Bibliotheksservice-Zentrum Baden-Württemberg

Aufwände BSZ und Erfahrungen:

- Mehraufwand: Erschließung, Parametrisierung, Crawling (Test und Produktion); automatisierte Datenübernahme für LZA in B-W
- Es entfallen: Kommunikation, Support, Schulung, Dokumentation
- Es entfallen: Administration Applikation + Kommunikations- und Präsentationsumgebung, lokale Datenverwaltung, lokale Speicherinfrastruktur
- Inanspruchnahme Helpdesk, Dokumentation, Schulungen, Webinare etc. von Archive-It durch BSZ
- Datenqualität: homogen und sehr gut

BSZ Bibliotheksservice-Zentrum Baden-Württemberg

„SaarDok“ 2018

- Zentraler Crawlingservice im BSZ
- Auftragscrawling nach schriftlicher Beauftragung
- rd. 165 Websites
- SULB Saarbrücken: Metadatenerfassung ZDB/SWB; Spiegelungserlaubnis; Beauftragung zu Crawling und Präsentation der Ressource; Erstabnahme Tests; QA für Produktion
- Recherche SWB, Präsentation Archive-It
- BSZ: Einrichten, Parametrisierung, Tests, produktiver Betrieb
- SULB: eingeschränkten Zugriff auf Workflow-Tool
- Aufwände erwartet wie Modell I + Schulung und Support QA

BSZ Bibliotheksservice-Zentrum Baden-Württemberg

- zentrale Webharvesting-Dienstleistung hat die Erwartungen erfüllt
- personelle und technische Ressourcen eingespart bzw. effizienter eingesetzt
- bessere Qualität in Crawling und Präsentation, zufriedene Kunden

- leistungsfähiges, nutzerfreundliches Workflow-Tool
- Inanspruchnahme Fremddienstleistung

Vielen Dank für Ihre Aufmerksamkeit!

Für Fragen wenden Sie sich gerne an

renate.hannemann@bsz-bw.de