

Fachverfahren Webarchivierung : Geschäftsmodelle und ihre Unterstützung durch die Software SWBcontent

Stefan Wolf, BSZ
nestor Praktikertag 2014
Hannover

Geschäftsmodell – was ist denn das?

Wikipedia:

*„Ein Geschäftsmodell (engl. business model) beschreibt die **logische Funktionsweise** eines Unternehmens und insbesondere die spezifische Art und Weise, mit der es Gewinne erwirtschaftet.*

Weder in der Wissenschaft noch in der Praxis gibt es eine allgemein akzeptierte Definition.

*Grundsätzlich soll die Beschreibung von Geschäftsmodellen helfen, die **Schlüsselfaktoren des Unternehmenserfolges** oder Misserfolges zu verstehen, zu analysieren und zu kommunizieren.“*

Inhalt

- Hintergrund und Entwicklung
 - Erfahrungsbericht
 - Personal
 - Software
 - Hardware
 - Leitfrage: was und wieviel braucht es denn?
- ≠ Kosten
- ≠ Preise

Eher nicht Inhalt

≠ allgemeine Kostenmodelle

- „Economic Sustainability Reference Model“
- 4C-Projekt: Collaboration to Clarify the Costs of Curation
- generischer Ansatz zur Ermittlung von Kostenfaktoren, statt konkret entstehende Kosten im Einzelfall zu beziffern

➤ Heute eher induktiv:

- welche Aufwände entstehen bzw. sind entstanden?
- BSB: Projekt Langzeitarchivierung von Websites durch Gedächtnisinstitutionen
- nestor AG Kosten

Schreiben der Direktion des BSZ an die Direktionen der Badischen und württembergischen Landesbibliothek

„In enger Kooperation soll ein Depot- und Portalsystem aufgebaut werden, dessen Ziel die langfristige Archivierung und Verfügbarkeit online verfügbarer Publikationen ist. In enger Abstimmung und Zusammenarbeit mit Der Deutschen Bibliothek wird eine redundante Datenhaltung vermieden. Dabei sollen austauschbare Beschreibungsdaten entstehen sowie nachnutzbare Werkzeuge für die Haltung, Verfügbarkeit und langfristige Sicherung und Benutzbarkeit der Objekte selbst. Die Nutzung wird frei im WWW bzw. - sofern lizenzrechtliche Gegebenheiten dies verlangen – nur beschränkt möglich sein.

Haupttätigkeitsfeld des BSZ wird die technische Unterstützung der Landesbibliotheken sein: die Bereitstellung ausreichender Server- und Speicherkapazitäten ist zu verbinden mit Verfahren, die eine rationelle Erschließung und Sammlung bzw. Ablieferung ermöglichen. Die Auswahl, Bestimmung und Erschließung der einzelnen zu archivierenden Objekte liegt besonders in der Verantwortung und Regie der beiden Landesbibliotheken.

Wir sind uns einig, dass eine genaue Bestimmung der einzusetzenden Personalkapazität und des Aufwandes an Sachmitteln noch nicht möglich ist. Ich habe schon heute die mit der Digitalen Bibliothek beschäftigten Mitarbeiter angewiesen, besondere Priorität auf unsere Kooperation zu legen. Gemeinsam werden wir **Mittel einwerben**, die den Aufbau dieses landeseigenen Depot- und Portalsystems ebenso wie die Teilnahme an der geplanten überregionalen Kooperationsplattform erlauben. Wir sollten rasch eine gemeinsame Arbeitsgruppe installieren, die dauerhaft die Einzelfragen klärt, aber auch einen ersten, pragmatisch gestalteten Einstieg realisiert.“

1. Ergebnis: Installation BOA



The screenshot shows the homepage of the Baden-Württembergisches Online-Archiv (BOA). The header features the BOA logo in red and white, with the text 'Baden-Württembergisches Online-Archiv' below it. To the right of the logo are navigation links: 'BADISCHE LANDESBIBLIOTHEK', 'LANDESARCHIV BADEN-WÜRTTEMBERG', and 'WÜRTTEMBERGISCHE LANDESBIBLIOTHEK', each with a right-pointing arrow icon. Further right are links for 'Home', 'Kontakt', and 'Hilfe'. The BSZ logo and a coat of arms are also visible in the top right corner.

BOA: Baden-Württembergisches Online-Archiv

Websites und Publikationen im Internet aus oder über Baden-Württemberg werden im "Baden-Württembergischen Online-Archiv" (BOA) gesammelt, erschlossen, der Öffentlichkeit zur Verfügung gestellt und für die Zukunft archiviert. Träger des Projekts sind die Badische Landesbibliothek in Karlsruhe, die Württembergische Landesbibliothek in Stuttgart, das Landesarchiv Baden-Württemberg und das Bibliotheksservice-Zentrum Baden-Württemberg.

Hier geht es zum Webseiten-Archiv des Landesarchivs:

- [BOA Landesarchiv](#)

Den Zugang zu den Dokumenten der Badischen Landesbibliothek und der Württembergischen Landesbibliothek finden Sie hier:

- [BOA Landesbibliotheken](#)

Viele Informationen, die früher in Form von Druckschriften in den Landesbibliotheken oder als Behördenschriftgut im Landesarchiv zugänglich gemacht wurden, sind heute nur noch über das Internet erhältlich. BOA stellt diese Dokumente stabil und dauerhaft zur Verfügung.

2. Ergebnis: Software SWBcontent

The image displays a stack of overlapping screenshots of various library and archival websites. The most prominent screenshot is for 'SWBregio: Digitales Webseitenarchiv'. It features a search interface with a search bar and a 'Suche starten' button. The text on the page includes: 'SWBregio: Digitales Webseitenarchiv', 'In SWBregio werden Webseiten von Städten, Kommunen und den mit ihnen verbundenen Körperschaften gesammelt, erschlossen, der Öffentlichkeit zur Verfügung gestellt und dauerhaft archiviert.', 'SWBregio ist eine Kooperation von Landesarchiv Baden-Württemberg und Bibliothekservice-Zentrum Baden-Württemberg. SWBregio steht offen für kommunale und regionale Anbieter von Webseiten, die ihre Internetauftritte archivieren möchten. Voraussetzung ist, dass durch die Institutionen, die Inhalte in SWBregio erstellen, die Rechte der Publizierenden gewahrt werden.', and 'Der Betrieb und die technische Sicherung der Inhalte liegen in den Händen des BSZ.'

Aber was braucht es dazu?

➤ **Zielsetzung**

➤ **Personal**

➤ **Software**

➤ **Hardware**

Zielsetzung

- **Beispiel: Materialauswahl / Einleitungstext BOA**
 - „Websites und Publikationen im Internet“
 - Auswahl? Bestimmung? Abgrenzung?
 - Verstehen Archivare und Bibliothekare das Gleiche?
- **Beispiel: Beschreibungsdaten**
 - Verzahnung mit übriger Beschreibung
 - Datenhaltung und –bereitstellung (Standalone bis Verbund – oder: die Parkschützer und der Juchtenkäfer)
- **Weitere Beispiele: Frequenz, Verfügbarkeit etc.**

Exkurs: Der Juchtenkäfer

Osmoderma eremita



Personal

- 12 Jahre Projektlaufzeit
 - „Verbrauch“:
 - 4 TechnikerInnen
 - 6 BibliothekarInnen
 - 1 über die ganzen Jahre
 - Dazu:
Management, IT-Administration etc.
 - Dazu:
 - Betrieb und Entwicklung:

Hardware

- **Ausreichend dimensioniert:**
 - Arbeitsspeicher und Speicherplatz
 - Test- und Produktionssysteme
- **laufend aktuelle und gegenseitig abhängig:**
 - Betriebssysteme und Software (z.B. Heritrix)
 - Unix- und Windowssysteme (Zeichensätze!)
 - Hardware beeinflusst Darstellung – authentische Archivierung!
- **„Verbrauch“: ca. 4 Produktionssysteme**
 - Umstellung der Anwendung und Umzug der Daten!

Software

- „Heritrix und Wayback“

Aber genügt das?

Aufgaben der Software

- Dateinamen und URL-Encoding
- URN-Bildung und –Registrierung
- Einfluß der Wayback auf die Darstellung
- Einfluß der Wayback auf das Layout
- Roter Balken
- URL-, Metadaten- und Volltextsuche
- Statistik: Mengen und Nutzung
- Logauswertung
- Automatische Benachrichtigungen
- Hashsummenbildung und –prüfung
- Freie und beschränkte Bereitstellung der Inhalte
(in SWBcontent im Wesentlichen gelöst)

Weitere Einzelaufgaben

- Javascripthaltige Seiten
- CSS-/PHP-Skripte, die sich aus Datenbanken bedienen
- Social Media
- Mobile Angebote
- Intranets
- Deduplikation

! Aufgaben: viele !

? Modelle/Angebote ?



Bundeskonzferenz der Kommunalarchive beim Deutschen Städtetag

Empfehlung: Speicherung von kommunalen Webseiten

„Eine Archivierung kommunal bedeutender privater Websites ist inhaltlich wünschenswert und im Rahmen der Übernahme von Sammlungsgut anzustreben“

„Mittlerweile besteht die Möglichkeit, über Fremddienstleister Webseiten archivieren zu lassen. Dieser Service wird vornehmlich von Bibliotheken angeboten, die an der vollständigen Erfassung und Archivierung digitaler Publikationen arbeiten. Ein empfehlenswertes Beispiel hierfür gibt das BSZ.“

„recht günstiges Preis-Leistungsverhältnis“

Betriebsmodelle

- **Eigenbetrieb**
 - z.B.: Siemens
 - **1 Installation für eine Institution**
 - Beispiel: Literatur im Netz
 - **1 Installation für mehrere Institutionen**
 - BOA: Kooperation
 - SWBregio: Verbund
- Weitere Modelle denkbar

Angebotsfaktoren (derzeit)

- Einrichtung
- Betrieb
 - Eigen- oder Auftragsharvesting
 - Softwarepflege
- Beratung/Betreuung
 - Webseitenanalyse, Logauswertung
- Speicherkosten
- (Auf Anfrage: Migration – Vorteil: Heritrix erzeugt WARC)

Gibt es ein Ergebnis?

- Aufwand nicht unterschätzen:
 - Eine Heritrix-Installation „unter“ der Schreibtischplatte genügt auf Dauer nicht.
- Keine Angst – sondern Mut:
 - Aufgabe ist lohnend
 - Technik steht zur Verfügung und ist den Kinderschuhen entwachsen
- Und es gibt Austauschforen
 - Zu den Softwarelösungen, im Rahmen AWW und nestor

Für Fragen stehe ich gerne zur Verfügung!

Stefan Wolf

stefan.wolf@bsz-bw.de

07531/883050

... und danke für die Aufmerksamkeit