



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Kultur BAK
Schweizerische Nationalbibliothek NB

Webarchiv Schweiz

Auswertung der Testphase 2008: Zusammenfassung

Autorin	Barbara Signori
Datum	25. September 2008
Version	1.0 (Arbeitsfassung)
Registratur	12 e-Helvetica

1	Inhaltsverzeichnis	
1	Inhaltsverzeichnis	2
2	Änderungen im Dokument	2
3	Zweck des Dokuments	2
4	Ausgangslage	2
5	Auswertung	3
5.1	Auftrag.....	3
5.2	Beteiligte	3
5.3	Statistik.....	4
5.4	FAQ.....	4
5.5	Offene Fragen.....	8
6	Weiteres Vorgehen	9

2 Änderungen im Dokument

Version	Datum	Änderung
1.0	25.09.2008	Ersterstellung

3 Zweck des Dokuments

Das Dokument enthält die Zusammenfassung der Auswertung der Statistik und des Feedbacks zum Ermitteln und Anmelden von landeskundlich relevanten Websites von 12 Kantonsbibliotheken (Appenzell Ausserrhoden, Bern, Glarus, Graubünden, Luzern, Nidwalden, Obwalden, Schaffhausen, Thurgau, Tessin Bellinzona und Lugano, Zürich), von 1 Staatsarchiv (Basel-Landschaft) und von 2 weiteren interessierten Institutionen (Sozialarchiv Zürich und Liechtensteinische Landesbibliothek).

4 Ausgangslage

Webarchiv Schweiz ist eine Sammlung von landeskundlich relevanten Websites („graue online Literatur der Schweiz“). Webarchiv Schweiz archiviert dauerhaft das geistige Online-Kulturgut der Kantone und der Schweiz und stellt dieses zur Verfügung. Für den Aufbau der Sammlung arbeiten die Schweizer Kantonsbibliotheken und die Schweizerische Nationalbibliothek sowie weitere schweizerische Spezialbibliotheken zusammen.

11 Kantonsbibliotheken (Basel-Stadt, Freiburg, Genf, Jura, Neuenburg, Solothurn, St. Gallen, Uri, Waadt, Wallis, Zug) beteiligen sich seit Mai 2008 an der Betriebsphase von Webarchiv Schweiz.

13 Kantonsbibliotheken (Appenzell Ausserrhoden, Appenzell Innerrhoden, Bern, Glarus, Graubünden, Luzern, Nidwalden, Obwalden, Schaffhausen, Thurgau, Tessin Bellinzona und Lugano, Zürich), 1 Staatsarchiv (Basel-Landschaft) und 2 weitere interessierte Institutionen (Sozialarchiv Zürich und Liechtensteinische Landesbibliothek) führen 2008 eine Testphase durch, um Webarchiv Schweiz aus einer praktischen Sicht kennen zu lernen. Dies ermöglicht ihnen den späteren Entscheid, Webarchiv Schweiz definitiv beizutreten.

Die Testinstitutionen haben am 23. und 30. April 2008 an einem Workshop zu Webarchiv Schweiz teilgenommen. Schwerpunkte des Workshops waren das Sammeln und Erschliessen von landeskundlich relevanten Websites.

5 Auswertung

5.1 Auftrag

Die Testinstitutionen haben am 15. Mai 2008 den Auftrag erhalten, landeskundlich relevante Websites zu ermitteln und anzumelden. Als Grundlage zur Ermittlung dienten die Richtlinien, die im Merkblatt Sammeln (Version 1.4) aufgeführt sind. Als Grundlage zur Anmeldung dienten die Richtlinien, die im Merkblatt Erschliessen (Version 1.2) aufgeführt sind. Unterstützung boten auch die Ausbildungsunterlagen.

Pro Testinstitution sollten insgesamt zwischen 10 bis 20 Websites ermittelt und angemeldet werden. Die Testdaten werden nach Abschluss der Testphase wieder gelöscht. Die Testphase endete am 30. Juni 2008. Daraufhin schickten die Testinstitutionen ihre Statistiken und Feedbacks.

In der NB wurden die Testdaten weiter verarbeitet. Bei allen Anmeldungen wurde überprüft, ob sie den Sammelkriterien des Merkblatts Sammeln entsprechen und ob die Erschliessung via Webformular korrekt erfolgt ist. Die Prüfung in der NB wurde ebenfalls dokumentiert.

Die Statistiken und Feedbacks wurden in der NB detailliert ausgewertet. Ziel war, Verbesserungsmöglichkeiten beim Ermitteln und Anmelden von landeskundlichen Websites zu identifizieren.

5.2 Beteiligte

Institution	Verwendete Abkürzung	Ansprechperson
Appenzell Ausserrhoden, Kantonsbibliothek	AR	Lipp, Patrick
Universitätsbibliothek Bern	BE	Bürger, Ulrike
Staatsarchiv Basel-Landschaft	BL	Quillet, René
Landesbibliothek des Kantons Glarus	GL	Bleichenbacher, Bettina
Kantonsbibliothek Graubünden	GR	Ursch, Antonia
Liechtensteinische Landesbibliothek	LI	Pinard, Lino
Zentral- und Hochschulbibliothek Luzern	LU	Mühlebach, Pius
Kantonsbibliothek Nidwalden	NW	Gut, Brigitte
Kantonsbibliothek Obwalden	OW	Hirter, Regula
Bibliotheken der Stadt Schaffhausen	SH	Giger, Dorothea
Schweizerisches Sozialarchiv	SOZARCH	Gubser, Susanne
Kantonsbibliothek Thurgau	TG	Gossweiler, Stephan
Biblioteca cantonale di Bellinzona	TI B	Dosi, Davide
Biblioteca cantonale di Lugano	TI L	Vögeli, Urs
Zentralbibliothek Zürich	ZH	Utzinger, Anita

Die Innerrhodische Kantonsbibliothek konnte die Tests nicht in der angegebenen Zeitspanne durchführen; sie wird die Tests später machen. Deshalb sind die Statistik sowie das Feedback von dieser Kantonsbibliothek nicht enthalten.

5.3 Statistik

Institution	Angemeldete Websites	Durchschnittliche Dauer für die Ermittlung	Durchschnittliche Dauer für die Anmeldung
AR	21	-	15 Min.
BE	13	7 Min.	25 Min.
BL	14	1 Min.	10 Min.
GL	11	-	33 Min.
GR	20	2,5 Min.	20 Min.
LI	20	19 Min.	5 Min.
LU	8	2 Min.	15 Min.
NW	12	3 Min.	2 Min.
OW	18	-	10 Min.
SH	20	8 Min.	15 Min.
SOZARCH	15	-	-
TG	16	1 Min.	13 Min.
TI B	11	-	-
TI L	8	1 Min.	8 Min.
ZH	20	27 Min.	35 Min.

Einigen Testinstitutionen fiel es schwer, die aufgewendete Zeit entweder der Ermittlung oder der Anmeldung zuzuordnen.

5.4 FAQ

Nachfolgend eine Zusammenstellung von ausgewählten Fragen mit den entsprechenden Antworten.

5.4.1 Ermitteln

Frage	Antwort
Im Kanton Zürich sind sehr viele Vereine und Firmen beheimatet (Hauptsitz), die sich „schweizerisch“ nennen und/oder von nationaler Bedeutung sind, die mit dem Kanton im landeskundlichen Sinn aber weiter nichts zu tun haben. Wer sammelt diese Websites?	Die Sammlung von Websites mit nationaler Bedeutung gehört in den Kompetenzbereich der NB. Bisher hat die NB ihre Sammeltätigkeit aber ausschliesslich auf Event-Harvestings (z.B. Wahlen 2007 und Euro 2008) beschränkt. Weiter werden durch die Ausweitung der Partnerinstitutionen von Webarchiv Schweiz (z.B. das Schweizerische Sozialarchiv) auch Websites mit nationaler Bedeutung zu bestimmten Fachbereichen gesammelt.
Gehören Lokalmedien wie Lokalzeitungen zum Sammelbereich?	Zeitungen und Zeitschriften werden in Webarchiv Schweiz nicht gesammelt. Sie sind Gegenstand eines anderen Pilotprojekts der NB. Bei Webarchiv Schweiz werden ganze Websites gesammelt und nicht einzelne Hefte oder Nummern von Zeitschriften.
Auf der Website der Hallen für Neue Kunst (www.modern-art.ch) wird nichts über Schaffhausen gesagt, trotzdem gehören sie zu Schaffhausen und prägen die Stadt/Region. Gehören solche Websites zum Sammelgebiet?	Wenn es sich um kulturelle Einrichtungen handelt, die sich im Kanton befinden und mit dem Kanton verbunden sind, dann gehören sie zum Sammelgebiet. Vorausgesetzt natürlich die Kantonsbibliothek erachtet die Website als archivierungswürdig.
In Schaffhausen gibt es die Firma Georg Fischer, welche im Kanton Thurgau das Klostersgut Para-	Dies liegt im Ermessen der Kantonsbibliothek. Im konkreten Beispiel würde die NB dem lokalen Sitz

dies (www.klostergutparadies.ch) und die Eisenbibliothek besitzt. Gilt hier der Firmensitz oder der lokale Sitz der Institution Klostergut Paradies, d.h. welche Kantonsbibliothek - Schaffhausen oder Thurgau - soll die Website vom Klostergut Paradies anmelden?	Vorrang geben. Wird eine Website von mehreren Kantonsbibliotheken angemeldet, werden die bibliografischen Aufnahmen in der NB manuell fusioniert und die Daten somit zusammengeführt.
Gibt es Formate, die nicht akzeptiert werden? Falls solche Formate innerhalb einer Website vorkommen, werden diese Teile dann nicht eingesammelt?	Beim Einsammeln von Websites gibt es keine Einschränkung bezüglich der Datenformate. Einschränkungen würden ja lediglich dazu führen, dass die Websites unvollständig gesammelt würden, weil ein Ausweichen auf alternative Formate nicht möglich ist. Allerdings führt die Art und Weise wie gewisse Daten angeboten werden trotzdem dazu, dass wegen technischer Limiten des Harvesting-Programms gewisse Files nicht eingesammelt werden können. Besonders betroffen davon sind Video- und Tondokumente. Das Problem ist aber nicht das Datenformat sondern das verwendete Übertragungsprotokoll.

5.4.2 Anmelden

Frage	Antwort
Was muss das eigene EDV-System leisten, damit der Zugriff zum Datenspeicher der NB klappt?	Die bibliografischen Aufnahmen können aus dem Bibliothekskatalog Helveticat der NB in den lokalen Bibliothekskatalog importiert werden. Diese Datenrückmeldung definieren wir in einer allfälligen Betriebsphase bilateral mit den Institutionen. Der Zugriff erfolgt immer über den eindeutigen Identifikator, die URN (Uniform Resource Name), die sich in der bibliografischen Aufnahme als anklickbarer Link befindet. Das lokale EDV-System muss als einzige Anforderung erbringen, dass den Benutzenden aktive Links zur Verfügung gestellt werden können. Geplant ist, ein eigenes Suchinterface für Webarchiv Schweiz über Helveticat zu erstellen. Ein Prototyp ist im Merkblatt Bereitstellen abgebildet. Ein Import der Daten in den lokalen Bibliothekskatalog ist somit nicht zwingend notwendig. In der NB wird das Zugriffssystem für die digitalen Sammlungen zurzeit entwickelt. Ziel ist, 2010 die archivierten Websites zugänglich zu machen.
Warum ist der Zeichensatz in der Bestätigungsmail nicht korrekt? Spielt das eine Rolle für die Verarbeitung?	Die Umlaute sind in der Bestätigungsmail falsch dargestellt. Das liegt an der Einstellung des Zeichensatzes. In der angehängten XML-Datei sind die Umlaute aber korrekt vorhanden. Für die Weiterverarbeitung der Anmeldungen ist nur die XML-Datei relevant.
Wie können Fehler bei der Anmeldung korrigiert werden?	Fehler sollen dank des Überprüfungsbildschirms möglichst vermieden werden. Eine spätere Korrektur durch die anmeldende Institution ist nicht möglich. Notfalls kann eine Anfrage an webarchiv-schweiz@nb.admin.ch geschickt werden. Dort wird entschieden, ob die Korrektur ma-

	nuell in der NB vorgenommen wird oder ob die anmeldende Institution die Website nochmals neu anmelden muss und die fehlerhafte Anmeldung gelöscht wird.
Die Website Schaffhauserland des Tourismus-Service hat mehrere Adressen. Eine Hauptadresse (www.schaffhauserland.ch) und Adressen, die auf die Hauptadresse verweisen (z.B. www.schaffhausen-tourismus.ch). Müssen alle Adressen vermerkt werden?	Es wird nur die Hauptadresse angegeben. Die Adressen, die nur Verweise sind, werden nicht angegeben.
Müssen im Feld Körperschaften auch Verweisungen angegeben werden?	Bei den Körperschaften wird nur die Hauptansetzung angegeben. Die Ansetzung erfolgt gemäss Regeln der anmeldenden Institution. Verweisungen können im lokalen Bibliothekskatalog nachträglich angefügt werden, wenn die Institution die bibliografischen Aufnahmen aus Helvetica in ihr lokales System importiert.
Können ausgewählte Websites auch abgemeldet werden?	Im Formular kann bei Sammelhäufigkeit angegeben werden, wie oft die Website eingesammelt werden soll. Wenn die anmeldende Institution die Website nach einer gewissen Zeit „abmelden“ will, d.h. sie soll nicht mehr regelmässig eingesammelt werden, kann dies an webarchiv-schweiz@nb.admin.ch gemeldet werden.
Auf der Website des Kantons Bern (www.be.ch) sind sämtliche Links zu politischen Gremien, Gemeinden, Ämtern, Schulen, Museen, Festivals usw. enthalten. Werden bei der Anmeldung dieser Website alle diese Bereiche mitgesammelt oder müssen sie einzeln angemeldet werden?	Es ist zu beachten, dass bei der Angabe des Domain Names nicht automatisch auch die Subdomains eingesammelt werden. Wenn also die Domäne www.be.ch angegeben wird, ist z.B. die Subdomain www.erz.be.ch von der Sammlung ausgeschlossen. Das Gleiche gilt für die Gemeindeseiten, die alle über eigene Adressen verfügen, z.B. www.aarberg.ch . Diese Websites müssen einzeln angemeldet werden.
Muss die Ansetzung der Körperschaften im Formular streng nach den Vorgaben der anmeldenden Institution erfolgen?	Wenn die bibliografischen Aufnahmen ins lokale Bibliothekssystem importiert werden sollen, ist es von Vorteil, wenn die Ansetzung (gilt für alle Autoritäten: Autoren, Körperschaften, Schlagwörter) gemäss Regeln der anmeldenden Institution erfolgt. Damit fallen keine nachträglichen Korrekturen im lokalen Bibliothekskatalog mehr an.
Welche Codes werden bei fehlendem Impressum vergeben?	Eine fehlende Urheberangabe ist eher selten und bisher nie vorgekommen. Bei Switch (https://nic.switch.ch/reg/indexView.action?request_locale=de) sind zumindest die Besitzer der URL registriert. Allerdings sollten von dort nur Körperschaften und nicht Privatpersonen übernommen werden.
Bei der Website der Pro Natura Ticino lässt sich keine Hauptadresse zuordnen; die einzelnen Rubriken (Attualità, Riserve usw.) verfügen über unterschiedliche Adressen. Wie kann die ganze Website angemeldet werden?	Wenn die Website Pro Natura Ticino mit der URL www.pronatura.ch/ti angemeldet wird, werden alle Rubriken eingesammelt. Die Tatsache, dass ein Verzeichnisnamen auf eine bestimmte Datei umgeleitet wird ist nicht aussergewöhnlich. Im Normalfall heisst die Datei index.htm(l). Auf dem Webserver lässt sich aber einrichten, dass statt

	index.htm halt novita.php aufgerufen wird, wenn nur ein Verzeichnisnamen aufgerufen wird.
Welche Unter-Sites sollen von der Archivierung ausgeschlossen werden und wie können diese ohne enormen Zeitaufwand eruiert werden?	Es ist von Vorteil, wenn die ganze Website angemeldet und eingesammelt wird. Ein aktives Suchen nach Unter-Sites, die ausgeschlossen werden sollen, ist nicht vorgesehen. Von technischer Seite her gesehen, machen v.a. Kalender, Datenbanken usw. Probleme. Trotzdem können Websites grundsätzlich als Ganzes angemeldet werden. Treten später Probleme beim Harvesting auf, wird das der Informatiker kontrollieren und beheben. Es ist denkbar, dass das Harvesting geändert oder gar abgebrochen werden muss. Eine Information an die anmeldende Institution ist vorgesehen.
Muss bei einem „Beitrag aus“ immer auch die übergeordnete Website angemeldet werden?	Die übergeordnete Website muss nicht angemeldet werden. Der Titel der übergeordneten Website muss im Feld „Beitrag aus“ aufgeführt werden. Dabei handelt es sich meistens um html-Seiten. PDF-Dokumente werden in der Regel als unabhängig betrachtet.
Die Website www.savognin.ch verfügt über zwei verschiedene Layouts, eines für den Sommer und eines für den Winter. Die meisten Informationen sind identisch, aber eigentlich sind es trotzdem zwei verschiedenen Seiten. Wie meldet man diese Seiten an?	Mit der Anmeldung der URL www.savognin.ch werden beide Teile (also Sommer und Winter) eingesammelt. Es genügt eine einzige Anmeldung. Denn die Verzeichnisstruktur unterscheidet nicht klar zwischen Sommer und Winter, deshalb kann der Harvester diese Unterscheidung nicht nachvollziehen. Wenn die Wörter Sommer und Winter in der bibliografischen Aufnahme vorkommen sollen, kann das z.B. durch die Angabe im Untertitel oder in der Titelvariante abgedeckt werden.
Muss bei der Beschlagwortung auf die anderen anmeldenden Institutionen Rücksicht genommen werden?	Das Schlagwortfeld ist von jeder Institution gemäss ihren gültigen Regeln für die Beschlagwortung auszufüllen. Eine Rücksichtnahme auf andere Institutionen ist nicht notwendig, da jede Institution nur die bibliografischen Aufnahmen zurück erhält, die sie auch selber angemeldet hat. In Helvetica werden die eingegebenen Schlagwörter als unkontrollierte Stichwörter aufgeführt. Eine weitere Beschlagwortung findet in der NB nicht statt.
Wieso wird kein Feld 245 \$c für den Urheber angelegt? Eine Eintragung müsste eigentlich in der bibliografischen Aufnahme verankert sein.	Bei den meisten Websites handelt es sich um integrierende Ressourcen. Die Körperschaft ist aus dem Titel oder dem Produzentennamen ersichtlich. Die Körperschaft wird nur eingetragen, wenn es sich um die Website der Körperschaft handelt und diese auch im Titel oder Produzentennamen vorkommt. Damit ist eine Verankerung in der bibliografischen Aufnahme gewährleistet. Persönliche Autoren werden nur bei statischen Ressourcen angegeben. Dabei wird das Feld 245 \$c automatisch im Hintergrund erstellt.

5.5 Offene Fragen

Nachfolgend sind offene Fragen aufgeführt, die aus den Feedbacks der Testinstitutionen abgeleitet werden konnten. Die Begleitgruppe von Webarchiv Schweiz prüft nun die offenen Fragen und setzt – wo möglich und angemessen – die Verbesserungen um.

5.5.1 Ermitteln

- „Ausländische“ Websites doch ausschliessen? Nur .ch auswählen inkl. .com, .org usw.?
- Bilder/Fotos von der Sammlung ausschliessen?
- Erklären welche Websites aus technischen Gründen nicht eingesammelt werden können inkl. Auflistung von Beispielen?
- Hilfe oder Vorlagen für ein „Muster-Sammelprofil“ erstellen? Richtlinien?
- Kriterien für die Beurteilung der technischen Eigenschaften von Websites festlegen?
- Kriterien für die Bewertung der Websites festlegen bzw. festlegen? Z.B. wie Deutschen Internetbibliothek?
- Muss der Begriff "nicht kommerziell und nicht kostenpflichtig" auf die Körperschaften der Websites ausgeweitet werden? Bisher wurde er so verstanden, dass die Websites an sich nicht kommerziell und nicht kostenpflichtig sein sollten, d.h. nicht passwortgeschützt usw. Es wurden aber Websites von lokalen Zeitungen, TV-Stationen, Verlagen (NZZ), Banken usw. angemeldet. Ist das problematisch?
- Punkt 4.2 im Merkblatt Sammeln ergänzen (z.B. Foto, Rundbrief, Filme, Blog)?

5.5.2 Anmelden

- Ausführliche Beispielsammlung mit vollständig ausgefüllten Formularen für jeden der drei Sammelbereiche im Merkblatt Erschliessen anhängen?
- Beispiele auszuschliessender Unterseiten/Dateien in den Hilfstexten aufführen? Z.B. Karten, die sich immer neu zusammensetzen; Datenbanken innerhalb der Website?
- Bessere Erklärung in den Hilfstexten zur Eingabe der URL, falls diese ändert? Wann soll die geänderte Form angegeben werden (bei Verweisen), wann soll das Anhängsel weggelassen werden usw.?
- Feld für Fussnoten anfügen (für Informationen, die keinem der bestehenden Felder zugeordnet werden kann)?
- Hilfstext beim Vorgehen nicht auffindbarer Produzenten ergänzen?
- Hilfstext zur Eingabe der URL (mit oder ohne /) besser erläutern bzw. prominenter hervorheben?
- Ist die Regelung mit dem Willkommensgruss zu kompliziert? Besser umschreiben oder ganz weglassen?
- Möglichkeit von mehr als einer Dewey-Vergabe?
- Problem mit Bestätigungsmail bei einzelnen Testinstitutionen – lokales Problem?
- Reichen 4 Sprachangaben nicht aus? Muss man das erhöhen?
- Sammelbereiche aus Formular entfernen – Deweyauswahl für alle?
- Sprachversionen in Hilfstext noch besser umschreiben, mit Beispielen?
- Weitere Schulung organisieren (auf Hilfstext hinweisen!) inkl. Hotline für Erschliessungsfragen aufbauen?
- Zusätzliches Feld für lokale Klassifikationen?
- Zwischenspeicherung des Formulars bei anmeldender Institution ermöglichen?

6 Weiteres Vorgehen

Nachfolgend die vorläufige Grobplanung für das weitere Vorgehen.

Was	Wer	Wann
In der Begleitgruppe Webarchiv Schweiz die offenen Fragen prüfen, besprechen und Verbesserungen umsetzen	Begleitgruppe Webarchiv Schweiz Mitarbeiter/innen e-Helvetica der NB	Von Oktober 2008 bis Mai 2009
Workshop zur Besprechung der Testphase durchführen, Erfahrungsaustausch und Diskussion	Technische Ansprechpersonen der Testinstitutionen Begleitgruppe Webarchiv Schweiz Mitarbeiter/innen e-Helvetica der NB	Im Januar 2009
Entscheid betr. Beteiligung am Betrieb von Webarchiv Schweiz fällen, Vereinbarung	Strategische Ansprechpersonen der Testinstitutionen	Im Februar/März 2009
Beitritt zu Webarchiv Schweiz	Testinstitutionen	Ab Mai 2009