



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Po Vogler – Realisierung Massnahme 6

Anforderungen an die Digitalisierung der geologischen Archive in Bund und Kantonen

Herausgeber
Bundesamt für Landestopografie swisstopo
Seftigenstrasse 264, Postfach
CH-3084 Wabern

Tel. +41 58 469 01 11
Fax +41 58 469 04 59
info@swisstopo.ch
www.swisstopo.ch

Inhaltsverzeichnis

1	Allgemeines	3
1.1	Sinn und Zweck.....	3
1.2	Datenaufbereitung in zwei Stufen	3
1.3	Seiten – Dokumente – Archiveinträge	3
1.4	Dateinamen.....	4
2	Scannen der Dokumente	5
2.1	Scanvorgang	5
2.2	Dateiformate.....	6
2.3	Qualität	6
2.4	Rahmenbedingungen.....	6
3	Aufbereitung und Qualitätssicherung	7
3.1	Aufbereitung	7
3.2	Qualitätssicherung	7
4	Weiterbearbeitung	7
5	Bei Fragen und Unklarheiten	7

1 Allgemeines

1.1 Sinn und Zweck

Eine Massnahme aus dem Postulat von Alt-Nationalrat Karl Vogler (Po Vogler) sieht vor, dass die geologischen Daten und Informationen von nationalem Interesse aus den dezentralen Archiven der Schweiz zusammengeführt und zentral verfügbar gemacht werden. Voraussetzung hierfür ist, dass die Bestände der kantonalen Archive in digitaler Form vorliegen. Wo dies noch nicht der Fall ist, müssen die analogen Archivbestände digitalisiert werden.

Für die Digitalisierung der geologischen Archive (nachfolgend Archive) in Bund und Kantonen (auf Wunsch auch Dritte) sind zwei grundsätzliche Szenarien denkbar:

1. swisstopo bietet den Archiven das Scannen der Dokumente als Service an. Die Archive ihrerseits stellen die Qualitätssicherung sowie die Zuweisung der Metadaten, Zugriffsrechte etc. sicher.
2. Die Archive stellen die Digitalisierung nach den Vorgaben von swisstopo selbst sicher (ggf. mit der Unterstützung eines externen Partners).

Unabhängig vom gewählten Szenario muss die Digitalisierung nach einheitlichen Vorgaben erfolgen. Diese werden im vorliegenden Dokument beschrieben.

Alle organisatorischen Fragen (Vorgehen, Finanzierung, Zeitplan etc.) sind nicht Bestandteil dieses Dokuments und werden separat behandelt (siehe Kontaktpersonen in Kapitel 5).

1.2 Datenaufbereitung in zwei Stufen

Bei der Digitalisierung von Archiveinträgen bestehen gegensätzliche Anforderungen. Einerseits sollen die Originaldokumente möglichst verlustfrei digitalisiert werden, was je nach Grösse der Vorlage und je nach Dateiformat zu sehr grossen Dateien führt (bis zu mehreren 100 MB). Andererseits sollen die digitalisierten Dokumente von Anwendern schnell heruntergeladen und möglichst an jedem Standard-PC oder ggf. mit Mobile Devices performant angezeigt werden können. Um beiden Ansprüchen gerecht zu werden, wurde entschieden, die Digitalisierung in zwei Stufen vorzunehmen:

1. Erstellen eines Rohformats/Grundformats zur Archivierung in optisch möglichst verlustfreier Qualität => JPEG-Format (.jpg-Dateien) mit 300 dpi Auflösung und maximaler Qualität (Einstellung: Stufe 12)
2. Erstellen eines speicherminimierten Gebrauchsformats für die Bereitstellung via Internet => PDF mit einer Pixelauflösung von 200 dpi und einer JPG-Qualitätsstufe von 10

Diese zweistufige Aufbereitung wird heute bereits im Bundesarchiv (BAR) erfolgreich umgesetzt. Auch bei swisstopo hat sich die zweistufige Datenaufbereitung bewährt. Zudem bietet das Verfahren jedem Archiv die Wahl, ob es beide oder nur eines der beiden erzeugten Formate bei sich aufbewahren will.

1.3 Seiten – Dokumente – Archiveinträge

Bei den hochauflösenden Archivkopien wird pro gescannte Seite eine einzelne JPG-Datei erzeugt. Für die Gebrauchskopien werden hingegen die Seiten analog den «physischen Dokumenten» zu PDF-Dokumenten zusammengeführt. Ein Archiveintrag (Dossier) kann mehrere Dokumente enthalten und jedes Dokument kann mehrere Seiten bzw. JPG-Dateien bündeln (vgl. Abbildung 1).

1.4 Dateinamen

Damit sich die Dateien einer Archiv-ID zuordnen lassen, müssen sie einheitlich bezeichnet werden:

JPG-Files:

<Betreiber>_<Archiv>_<Archiv-ID>_<Dokument-Nr.>_<Nr. der Seite im Dokument>.jpg

Beispiel 1: CH_LG_32706_01_015.jpg

Beispiel 2: AG_ATB_1023_01_002.jpg

PDF-Files:

<Betreiber>_<Archiv>_<Archiv-ID>_<Dokument-Nr.>.pdf

Beispiel 1: CH_LG_32706_01.pdf

Beispiel 2: AG_ATB_1023_01.pdf

Die Dokument-Nummer ist immer 2-stellig. Die Seitennummer bei den JPG-Dateien ist immer 3-stellig. Die zu verwendenden Bezeichnungen der Betreiber befinden sich im Anhang.

Diese Dateibezeichnung wird auch auf digi.swissgeol.ch für die automatische Zuweisung der PDF-Dateien zu den Metadaten benötigt. Es ist deshalb wichtig, dass alle PDF-Dateien von nationalem Interesse, die an swisstopo transferiert werden sollen, nach dieser Nomenklatur bezeichnet werden.

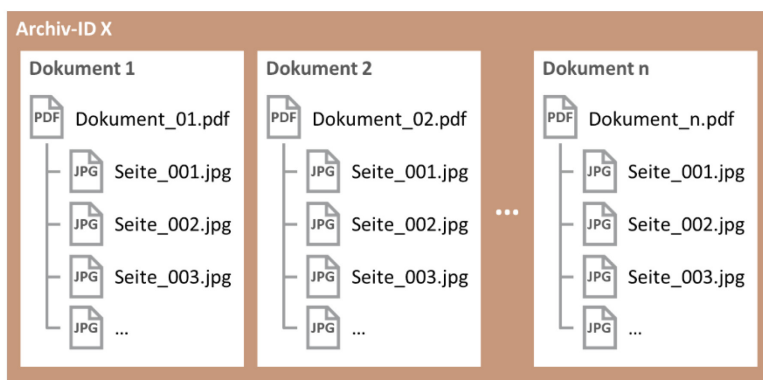


Abbildung 1. Struktur eines Archiveintrags

2 Scannen der Dokumente

2.1 Scanvorgang

Archivinhalt, die ganz oder teilweise in Papierform vorliegen, müssen gescannt werden. Die grundsätzlichen Arbeitsschritte hierzu zeigt Abbildung 2.

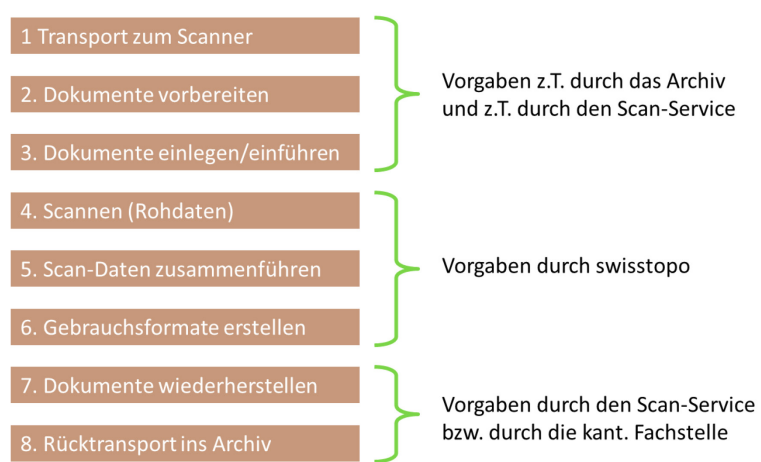


Abbildung 2. Ablauf beim Scannen analoger Archivinhalte

1. Die Dossiers müssen aus dem Archiv geholt und zum Scanner transportiert werden.
2. Jedes Dokument muss nach Seitenformat zerlegt und sortiert werden. Gebundene Dokumente müssen aufgetrennt werden.
3. Pro Format müssen die Dokumente in den Scanner eingelegt werden. Für grosse Dokumente wird ein Grossformatscanner benötigt.
4. Der eigentliche Scan-Vorgang wird durch das jeweilige Scanner-Modell bestimmt. Die Rohdaten werden als JPG-Dateien abgelegt.
5. Dokumente, die aufgrund unterschiedlicher Seitenformate aufgetrennt und auf unterschiedlichen Scannern eingescannt werden, müssen anschliessend korrekt zusammengeführt werden (korrekte Nummerierung und Ablage der Dateien).
6. Für die Weiterverwendung durch die Archive werden zusätzlich zu den Archivdateien «Gebrauchskopien» im PDF-Format mit geringerer Dateigrösse erstellt (kann mittels Skript automatisiert erfolgen).
7. Die aufgetrennten Papierseiten müssen wieder zu den ursprünglichen Dokumenten zusammengeführt und ggf. wieder gebunden werden (nach Vorgabe des jeweiligen Archivs).
8. Die fertig gescannten Dokumente müssen zurück ins Archiv transportiert und korrekt abgelegt werden.

swisstopo bietet jedem Kanton das Scannen der analogen Archive als Service an. Auf ausdrücklichen Wunsch kann das Scannen auch durch die Archive selbst sichergestellt werden (ggf. mit Unterstützung eines externen Partners). Dabei sind jedoch die Vorgaben in den Schritten 4, 5 und 6 gemäss Abbildung 2 zwingend einzuhalten.

2.2 Dateiformate

Rohformat	JPEG (maximale Qualität)
Gebrauchsformat	PDF Legacy, 200 dpi für alle Bilder über 200 dpi, Qualität hoch, bikubische Neuberechnung (für Farb- und Graustufenbilder), ohne OCR.

2.3 Qualität

Rohformat	- 300 dpi für Dokumente - 300 dpi für geologische Profile, Seismische Linien oder petrophysikalische Logs - 3'000 dpi für Mikrofichen - 3'000 dpi für Dias
Gebrauchsformat	- 200 dpi
Farbauflösung	- Farbig (RGB) für alle Dokumente

2.4 Rahmenbedingungen

Beim Scannen von Dokumenten im Format A3 und grösser ist folgendes besonders zu beachten:

- Gefaltete Dokumente müssen soweit geglättet werden, dass die Falten beim Scannen möglichst nicht mehr sichtbar sind.
- Grosse Dokumente müssen immer am Stück und möglichst verzerrungsfrei gescannt werden. Aus mehreren Einzelbildern zusammengesetzte Dokumente sind qualitativ unzureichend und deswegen nicht akzeptabel.
- Es muss sichergestellt werden, dass die Papiervorlage immer gerade eingezogen wird. Text- und Bildränder müssen immer parallel zum Seitenrand sein.

3 **Aufbereitung und Qualitätssicherung**

3.1 **Aufbereitung**

Die Dateiaufbereitung erfolgt nach den Grundsätzen in den Kapiteln 1.2 und Kapiteln 1.3. Einträge aus Archiven, die bereits digitalisiert und mit Metadaten versehen sind, müssen nicht nochmals neu gescannt werden.

3.2 **Qualitätssicherung**

Im Anschluss an die Digitalisierung und vor dem Upload auf digi.swissgeol.ch sind die gescannten Dokumente auf folgende Aspekte zu hin prüfen:

- Sind zu jedem Archiveintrag alle Dokumente vorhanden?
- Sind alle Dokumente nur einmalig vorhanden (keine Duplikate)?
- Sind innerhalb der Dokumente alle Seiten vorhanden (auch bei doppelseitigen Vorlagen)?
- Sind die Inhalte und die einzelnen Dokumente in der richtigen Reihenfolge angeordnet?
- Sind alle Inhalte in der richtigen Lage (keine gedrehten Inhalte)?

4 **Weiterbearbeitung**

Wie eingangs erwähnt wurde, müssen nach dem Digitalisieren der analogen Archivbestände die PDF-Dateien zusammen mit den entsprechenden Metadaten in die Webapplikation «digi.swissgeol.ch» hochgeladen werden. Dort können die Daten bereinigt, die PDF-Files mittels OCR lesbar gemacht und die Daten von nationalem Interesse zu swisstopo kopiert werden.

Erläuterungen zu den weiteren Prozessschritten stehen Ihnen auf digi.swissgeol.ch zur Verfügung. Als Archivbetreiber bzw. Archivadministrator eines kantonalen Archivs erhalten Sie per E-Mail eine Einladung mit den entsprechenden Zugangsinformationen.

5 **Bei Fragen und Unklarheiten**

Sind Vorgaben unklar oder fehlen Ihnen Informationen? Gerne stehen Ihnen Michael Gysi und Nils Oesterling zur Verfügung:

Michael Gysi	Prozessleiter Datenmanagement	Bundesamt für Landestopografie swisstopo Landesgeologie Seftigenstrasse 264 3084 Wabern michael.gysi@swisstopo.ch +41 58 465 82 28
Nils Oesterling	Projektleiter Dateninfrastruktur	Bundesamt für Landestopografie swisstopo Landesgeologie Seftigenstrasse 264 3084 Wabern nils.oesterling@swisstopo.ch +41 58 469 05 24

Anhang

Liste der Archivbetreiber für Bund und Kantone

Für die Dateibezeichnungen der kantonalen Archive sowie der Archive des Bundes sind zwingend die folgenden Abkürzungen im Dateinamen zu verwenden:

CH	Bund
AG	Aargau
AI	Appenzell Innerrhoden
AR	Appenzell Ausserrhoden
BE	Bern
BL	Basel-Landschaft
BS	Basel-Stadt
FR	Fribourg
GE	Genf
GL	Glarus
JU	Jura
LU	Luzern
NE	Neuenburg
NW	Nidwalden
OW	Obwalden
SG	St. Gallen
SH	Schaffhausen
SO	Solothurn
SZ	Schwyz
TG	Thurgau
TI	Tessin
UR	Uri
VD	Waadt
VS	Wallis
ZG	Zug
ZH	Zürich

Archivbetreiber für Dritte

Archive von Dritten (z.B. Universitäten, SBB, Private etc.) verwenden in ihren Dateinamen jeweils an erster Stelle die vollständige Bezeichnung des Archivbetreibers.

Beispiele: SBB_Archiv-x_1234_01_003.jpg
 UniBern_Archiv-y_54321_03.pdf
 Name-der-Unternehmung_Archiv-z_897431_03.pdf

Bei Fragen und Unklarheiten zu den Bezeichnungen wenden Sie sich bitte an die Kontaktpersonen in Kapitel 5.