

# Ablieferungsprozess Handelsregister

## Versionen

2006-05-11	Version 0	
2006-06-21	Version 1	Ergänzung mit Metadaten gemäss Zusammenstellung der AG Erschliessung (vgl. Kapitel 3.1.1 Attribute). Entsprechend wurde der zugehörige TODO-Eintrag geändert.
2006-07-04	Version 2	Kapitel 3.2.1 Spaltendelimiter und -titel korrigiert.
2006-08-02	Version 3	Überarbeitungen aufgrund der Ergebnisse der Sitzung AG Übernahme vom 6. Juli 2006 in Luzern

## Todo

Kapitel	Todo	Wer	Termin
3.1	Liste der Attribute zu den einzelnen Firmen für die Ablieferungsliste zur Diskussion stellen	Arbeitsgruppe	21. Juni 2006
3.1	Dateiname der Ablieferungsliste	Arbeitsgruppe	6. Juli 2006
1.1.1	Entscheid über Auswahl der Belegakten (integrale Übernahme oder Aufstellung der Kriterien für die Auswahl)	Durchführende Staatsarchive	bis 6. Juli 2006
2.2	Zusammenstellung der Metadaten	Jeder für sich, anschl. AG	bis 6. Juli 2006

## Inhaltsverzeichnis

<b>1 Tätigkeiten des Handelsregisteramtes.....</b>	<b>2</b>
1.1 Datenextraktion aus der Handelsregisteranwendung.....	2
1.1.1 Datenorganisation.....	2
1.1.2 Metadaten.....	2
1.2 Generierung der Prüfsummen für die Integritätsprüfung.....	2
1.3 Verpacken der Dateien.....	2
<b>2 Tätigkeiten des Archivs.....</b>	<b>3</b>
2.1 Überprüfen der Integrität, Vollständigkeit und Konformität der erhaltenen Daten.....	3
2.2 Zusammenstellen der Metadaten.....	3
2.2.1 Behältnis Metadaten (optional).....	3
2.2.2 Deskriptive Metadaten.....	3
2.2.3 Technische Metadaten.....	3
2.2.4 Administrative Metadaten.....	3
2.3 Generieren des Archivpakets.....	4
<b>3 Anhang.....</b>	<b>4</b>
3.1 Ablieferungsliste.....	4

3.1.1 Format.....	4
3.2 Checksummenliste.....	4
3.2.1 Format.....	5

## 1 Tätigkeiten des Handelsregisteramtes

### 1.1 Datenextraktion aus der Handelsregisteranwendung

Das Handelsregisteramt ist für die anforderungsgerechte Extraktion der Handelsregisterdaten verantwortlich. Die Anforderungen enthalten Angaben zu den zu extrahierenden Daten und der Form, wie diese an das Archiv abzuliefern sind.

#### 1.1.1 Datenorganisation

Die an das Archiv abzuliefernden Daten setzen sich wie folgt zusammen:

- ✓ Liste der gelöschten Firmen. Eine genaue Beschreibung der Liste findet sich im Kapitel 3.1 im Anhang.
- ✓ Belegakten zu den gelöschten Firmen (integral oder Auswahl). Die Belegakten sind in den Formaten TIFF und/oder PDF abgelegt. Sie können in dieser Form an das Archiv übergeben werden. Bevorzugt wird das PDF-Format, da dieses auch Volltextrecherchen zulässt. Da die Dateinamen der Belegakten eindeutig und (beinahe) selbsterklärend sind, können auch diese beibehalten werden. Es empfiehlt sich eine Ablieferung auf CD oder DVD oder aber via FTP.
- ✓ Dokumentation zur Anwendung (kann digital oder auf Papier erfolgen). Möglicherweise müssen wir damit warten, bis eine Dokumentation des Herstellers verfügbar wird. Zur Dokumentation zählen die rechtlichen Grundlagen, auf denen das Handelsregister jeweils basiert und weitere kontextbezogene Informationen.
- ✓ Codelisten und ggf. Abkürzungsverzeichnis soweit für das Verständnis der Daten notwendig. Diese müssen für den Benutzer lesbar sein.
- ✓ Liste der von den Belegaktendokumenten generierten Hashcodes. Zum Format der Liste siehe Kapitel 3.2 im Anhang.

#### 1.1.2 Metadaten

Zusätzlich zu den extrahierten Daten müssen folgende Metadaten mitgeliefert werden:

##### Deskriptive Metadaten

- ✓ Provenienzstelle
- ✓ Zeitraum, den die Extraktion umfasst (also Gründungsdatum der ältesten der gelöschten Firmen und jüngstes Löschedatum)
- ✓ Umfang (in MB)

##### Administrative Metadaten

- ✓ Datum und Zeitpunkt der Extraktion
- ✓ Die benutzten Abfragen (z.B. als SQL-Abfragen, XQuery o.ä.)
- ✓ Anzahl der abgelieferten Dateien
- ✓ Bezeichnung der Quellapplikation
- ✓ Version der Quellapplikation

Die Metadaten müssen ebenfalls in elektronischer Form vorliegen.

## 1.2 Generierung der Prüfsummen für die Integritätsprüfung

Von jeder Datei muss die Prüfsumme generiert werden. Als Algorithmus soll vorerst md5 verwendet werden. Die Prüfsummen werden in einer separaten Datei (vgl. dazu Kapitel 1.1.1 und Kapitel 3.2) abgelegt, die mit den übrigen Daten an das Archiv übergeben wird.

## 1.3 Verpacken der Dateien

Bei Bedarf können die alle Dateien zusammen in eine ZIP-Datei verpackt an das Archiv abgeliefert werden. Als Packmethode ist diesfalls Deflate zu benutzen. Das Zusammenpacken und komprimieren der Daten ist nicht notwendig.

# 2 Tätigkeiten des Archivs

Das Archiv übernimmt vom Handelsregisteramt die Daten. Die Ablieferungsart spielt keine Rolle, d.h. die Datei kann auf einem Datenträger oder über das Netz (z.B. FTP) geliefert werden.

## 2.1 Überprüfen der Integrität, Vollständigkeit und Konformität der erhaltenen Daten

In einem ersten Schritt sind die erhaltenen Daten zu überprüfen. Es sind folgende Prüfmassnahmen vorgesehen:

- ✓ Lesbarkeit aller Dateien überprüfen (können die Dateien in einem Editor geöffnet werden?). Abhängig von der Anzahl der Dateien und dem Prüfverfahren kann auch stichprobenmässig geprüft werden.
- ✓ Prüfsummencheck der Datendateien
- ✓ Vollständigkeit der Daten überprüfen (Anzahl Dateien, Anzahl Datensätze pro Datei)
- ✓ Angaben in der Metadatenfile überprüfen

Nach erfolgreicher Prüfung wird dem Handelsregisteramt eine Empfangsbestätigung übergeben, welche die Richtigkeit und Übereinstimmung der gelieferten Daten mit den mitgelieferten Metadaten bestätigt.

## 2.2 Zusammenstellen der Metadaten

Die Metadaten werden aus den vom Handelsregisteramt enthaltenen und zusätzlichen Metadaten zusammengestellt. Sie bestehen aus den deskriptiven Metadaten (das sind die Metadaten, die für die Verzeichnung vorgesehen sind), aus den technischen Metadaten (welche die Daten solchermassen beschreiben, dass sie wieder dargestellt werden können) und aus den administrativen Metadaten (welche vorgenommene Veränderungen, durchgeführte Migrationen etc. dokumentieren).

Die Metadaten müssen so zusammengestellt sein, dass sie auch von einem automatischen Prozess gelesen und interpretiert werden können. Daher eignet sich die bewährte Form der XML-Datei hervorragend. Im Rahmen des Pilotprojekts wird ein minimaler Satz verwendet, dem, nachdem er sich bewährt hat, je nach Anforderung weitere Metadaten hinzugefügt werden können.

### 2.2.1 Behältnis Metadaten (optional)

Dazu gehören eine Liste der im Behältnis zusammengefassten Dateien, die ihnen zugehörigen Prüfsummen. Für das Pilotprojekt kann ggf. auf ein spezielles Behältnis verzichtet werden. Dementsprechend sind die Metadaten anzupassen.

Zu einem späteren Zeitpunkt gehören hierher ebenfalls Angaben, welcher Typ von Behältnis vorliegt. Dies ist v.a. In Bezug auf die Möglichkeit der automatischen Aufbereitung der Daten für Benutzer von Bedeutung. Aufgrund dieser Angaben könnte ein entsprechendes Programm die benötigten Daten extrahieren, benutzergerecht aufbereiten und dem Benutzer in der gewünschten Form präsentieren.

## 2.2.2 Deskriptive Metadaten

Zu den deskriptiven Metadaten zählen z.B. die Signatur, Titel, Provenienzstelle, Laufzeit, Umfang und Verzeichnungsstufe. Nach Bedarf, kommen weitere, von der AG Erschliessung zu definierende Angaben, dazu.

## 2.2.3 Technische Metadaten (optional)

Zu den technischen Metadaten zählen Angaben zu den (vgl. dazu auch Kapitel 1.1.1) Dateieigenschaften, resp. Dateiformat der einzelnen Dateien im Behältnis. Diese werden beispielsweise bei einer automatisch durchgeführten Migration zur Feststellung des aktuellen Dateiformats benötigt.

## 2.2.4 Administrative Metadaten

Die administrativen Metadaten halten Veränderungen fest, die bei der Übernahme, bei Migrationen und ähnlichen Prozessen durchgeführt werden. Es handelt sich primär um protokollierung solcher Veränderungen mit mindestens dem Zeitpunkt, an dem die Veränderung durchgeführt worden ist, der Art der Veränderung und durch wen (Person oder Programm) sie durchgeführt worden ist.

Die Metadaten werden in eine (oder mehrere) Metadateien, die idealerweise ein XML-Format aufweisen, zusammengefasst. Ein entsprechenden Entwurf muss noch ausgearbeitet werden.

Nicht unbedingt als Metadaten kann die Benutzerdokumentation bezeichnet werden. Diese kann aus beliebig vielen Dateien bestehen, die mit den Primärdaten dem Benutzer mitüberlassen werden. Zu diesem Zweck erhalten sie in den Behältnismetadaten einen speziellen Hinweis.

## 2.3 Generieren des Archivpakets

Das Archivpaket wird in einer geeigneten Form an die AG Aufbewahrung übergeben. Für das Pilotprojekt ist die Behältnisform noch nicht geregelt. Zu einem späteren Zeitpunkt ist die Wahl einer bewährten Behältnisform empfehlenswert. Derzeit eignet sich das ZIP-Format. Folgende Dateien gehören zum Paket, das an die AG Aufbewahrung geliefert wird:

- ✓ Die Liste der gelöschten Firmen
- ✓ Die Belegakten in Form von PDF- (oder TIFF-) Dateien
- ✓ Die Metadatei (entweder ein Datei mit allen Metadaten (metadata.xml) oder mehrere Dateien, wobei ich erstere Lösung (eine Datei) aufgrund der leichteren Verwendbarkeit bevorzuge.
- ✓ Die Dokumentationen (alle Dateien, die beschreibenden Charakter haben und nicht durch ein Programm gelesen werden müssen; also z.B. die Benutzerdokumentation.
- ✓ Gegebenenfalls die Codelisten und Abkürzungsverzeichnisse
- ✓ Die Liste mit den Prüfsummen

Als Kompression muss Deflate verwendet werden, da diese Form am verbreitetsten ist und keinen rechtlichen Restriktionen unterliegt.

Als Name für das Archivpaket könnte die Signatur verwendet werden, sofern diese nicht gegen die Regeln des Repositoriums verstösst (in unserem Fall das Dateisystem des Betriebssystems).

## 3 Anhang

### 3.1 Ablieferungsliste

Form und Aufbau der Ablieferungsliste ist noch zu definieren. Als Vorlage können vorerst die Ablieferungsliste des Kantons Luzern (ein Beispiel habe ich Euch, zusammen mit einem Registerauszug,

seinerzeit per Email zugeschickt) und die vom Staatsarchiv Zürich erhaltenen Unterlagen (InformationDatenbankHRAJanuar2004.doc und loesch.txt) dienen. Die Ablieferungsliste soll in elektronischer Form abgeliefert werden. Mindestsatz (resp. vorgeschriebene Felder):

### 3.1.1 Attribute

Firmennummer
Firmenname
Firmensitz
Eröffnungsdatum
Löschungsdatum

### 3.1.2 Format

Dateiformat	Reines Textformat (MIME-Typ: text/tab-separated-values <sup>1</sup> )
Zeichensatz	UTF-8
Spaltendelimiter	TAB-delimited (HT, resp. chr(9))
Zeilendelimiter	CR+LF oder CR oder LF
Spaltenbeschriftung	Erste Zeile

## 3.2 Checksummenliste

Die Checksummenliste wird für die Verifikation der Integrität der Belegaktendateien benötigt. Sie wird vom Handelsregisteramt generiert. Als Algorithmus wurde md5<sup>2</sup> gewählt. Diese Liste hat folgenden Aufbau (das Beispiel wurde mit dem Programm MD5summer generiert):

```
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 15.06.2006 08:26:25

f87c29ebe2821ea4d778d4abcfafb27b *filemaker pro 6.txt
ecd7b7592725a9de0a54b0846a15161e *galserve v6.cmd
707f732e15d815d9d720acd6bf17c059 *galserve v8.cmd
cd17bdb95861b5b6c5272ba175e11cc8 *Konfiguration NetOp.doc
b5c7bc4ed948df0c86e1598694f6393f *krankengeschichte.odt
8331bb11dd53c75024569cbf36d6bc3b *Liebe Jennifer.doc
2b2281f39be38f0aec9a0fd133b122fc *nicola.odt
...
```

Jede gültige Spalte beginnt mit der generierten Prüfsumme, gefolgt von einem Leerschlag. Anschliessend folgt der Name der Datei, aus der die Prüfsumme generiert wurde, beginnend mit einem Asterisk.

### 3.2.1 Format

Dateiformat	Reines Textformat (MIME-Typ: text/plain <sup>3</sup> )
Zeichensatz	UTF-8
Spaltendelimiter	Leerschlag + *

<sup>1</sup> Entnommen aus: <http://de.selfhtml.org/diverses/mimetypen.htm#uebersicht>

<sup>2</sup> Es gibt eine grosse Zahl von Public Domain Programmen, die für die Hashcode-Generierung geschrieben worden sind. Eine kleine Auswahl:  
<http://www.md5summer.org/>  
<http://gnuwin32.sourceforge.net/packages/textutils.htm>

<sup>3</sup> Vgl. Fussnote 1.

2006-07-04

Dateiformat	Reines Textformat (MIME-Typ: text/plain)
Zeilendelimiter	CR+LF oder CR oder LF
Spaltenbeschriftung	Keine