



Anleitung

Von Kansy Lambert

Betrifft Anforderungen an Transkriptionen von Archivgut

Datum/Vers. 11.4.2012 / 1-0

Status: Definitive Fassung

Dok.-ID 983846

Dossier Digitalisierung, Grundlagenpapiere

1. Einleitung

Das Staatsarchiv Basel-Stadt digitalisiert Archivgut zum Zwecke der Bestandsschonung und der Verbesserung der Zugänglichkeit. In der Regel beschränkt sich die Digitalisierung durch das Staatsarchiv auf die Erzeugung und Bereitstellung von digitalen Abbildern des analogen Materials. Eine Erstellung digitaler Volltexte erfolgt nicht.

Bei der Nutzung von Archivgut entstehen durch Arbeit der BenutzerInnen in zahlreichen Fällen Transkriptionen von Archivgut. Diese können die Image-Digitalisate in zweierlei Hinsicht ergänzen: zum Einen wird durch eine Integration der Transkripte in den Volltextindex des Archivinformationssystems die Auffindbarkeit von Informationen bis auf Stufe des Einzeldokuments über die Volltextsuche ermöglicht. Zum Anderen kann durch das Vorliegen eines Transkripts BenutzerInnen, die keine Fertigkeiten zur Lektüre von Handschriften besitzen, das Verständnis des Archivguts überhaupt ermöglicht werden.

2. Zielsetzung

Die vorliegenden Anforderungen an Transkriptionen und Editionen analogen Archivguts in digitaler Form stellen sicher, dass die Integration in die Retrievalinstrumente des Archivinformationssystems und die Einbettung in die Metadaten-datei des Digitalisats möglich ist.

3. Grundsätze

- Die nachstehenden Richtlinien verstehen sich als Mindestanforderungen, die bei einzelnen Transkriptions- und Editions-vorhaben erweitert werden können.
- Bei der Erstellung der Transkriptionen sind die allgemeinen Anforderungen wissenschaftlicher Forschung – Genauigkeit, Verlässlichkeit, Verifizierbarkeit, Unparteilichkeit und Unabhängigkeit – zu erfüllen.
- Die Richtlinien gelten für die Transkription handschriftlicher Texte
- Die Volltextfassung gedruckter Texte folgt konsequent der Druckvorlage. Die Erfassung von in Fraktur gesetzten Drucken folgt konsequent der Druckvorlage. Dabei sind die buchstabenformen zu normalisieren.
- Transkriptionen werden, sofern im StABS ein Digitalisat der Archivalien vorhanden ist, nicht nur in das Archivinformationssystem integriert, sondern auch in die Metadaten-datei des Digitalisats integriert.
- Ein Digitalisat besteht aus folgenden Komponenten:
 - Pro Seite der Archivalie eine Bilddatei
 - Für die Archivalie eine XML-Metadaten-datei nach dem METS-Standard.

In die Metadatei wird das Transkript als XML-Struktur integriert.

4. Nutzungsrechte des StABS an der Transkription

Damit die angestrebten Ziele auch realisiert werden können, ist neben der Einhaltung der nachstehenden Editions- und Transkriptionsregeln seitens der Herausgeber auch die Übertragung von Nutzungsrechten an das Staatsarchiv notwendig. Das Staatsarchiv kann Transkriptionen sowohl im Archivinformationssystem als auch in den Digitalisaten nachnutzen, wenn ihm diese unter einer [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 3.0 Schweiz Lizenz](#)¹ überlassen werden.

Die wesentlichen Bestimmungen dieser Lizenz lauten:

Das Werk bzw. der Inhalt darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden und es dürfen Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes erstellt werden, solange folgende Bedingungen eingehalten werden:

- Namensnennung — Der Name des Autors/Rechteinhabers ist in der von ihm festgelegten Weise zu nennen.
- Keine kommerzielle Nutzung — Dieses Werk bzw. dieser Inhalt darf nicht für kommerzielle Zwecke verwendet werden.
- Weitergabe unter gleichen Bedingungen — Wenn das lizenzierte Werk bzw. den lizenzierten Inhalt bearbeitet oder in anderer Weise erkennbar als Grundlage für eigenes Schaffen verwendet wird, dürfen die daraufhin neu entstandenen Werke bzw. Inhalte nur unter Verwendung von Lizenzbedingungen weitergeben, die mit denen dieses Lizenzvertrages identisch oder vergleichbar sind.

5. Anforderungen an die Übergabe der Transkription

Zusammen mit den unter Ziffer 5. genannten Objekten ist bei der Übergabe an das Staatsarchiv auch eine knappe Beschreibung der Transkription notwendig, die folgende Angaben enthält:

- Name, Adresse und Emailadresse des Herausgebers,
- Angabe des Kontextes, in dem die Transkription entstanden ist,
- Erklärung des Herausgebers, dass er die Transkription dem Staatsarchiv unter einer [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 3.0 Schweiz Lizenz](#) zur Nutzung überlässt.
- Dateinamen und Dateiformat der gelieferten Objekte,
- die Archivsignatur der transkribierten Archivalie,
- ggf. Angabe der URL des verwendeten Digitalisats des Staatsarchivs,
- Nennung der bei der Erstellung der Transkription verwendeten Anwendungen (inkl. Versionsangabe) und des Betriebssystems,
- Angabe, ob die Unicode-Text-Fassung Zeilenumbrüche enthält und wenn ja, wie diese markiert sind (siehe Ziffer 5 Pkt. 7),
- Hinweis, ob die Transkriptionsregeln unter Ziffer 7, angepasst worden sind,
- ggf. Nennung der Abweichungen von den Transkriptionsregeln oder Angabe der Online-Ressource, falls ein anderes Regelwerk verwendet wurde.
- ggf. Angabe, welche Anwendungen und Regelwerke/Style Sheets bei der Erstellung von XML-Dateien gemäss dem TEI- oder ALTO-Standard² verwendet wurden.

¹ vgl. <http://creativecommons.org/licenses/by-nc-sa/3.0/ch/> [besucht am 11.4.2012].

² vgl. zum TEI-Standard der Text Encoding Initiative (<http://www.tei-c.org>): "The TEI *Guidelines for Electronic Text Encoding and Interchange* define and document a markup language for representing the structural, rendition, and conceptual features of texts. They focus (though not exclusively) on the encoding of documents in the humanities and social sciences, and in particular on the representation of primary source materials for research and analysis. These guidelines are expressed as a modular, extensible XML schema, accompanied by detailed docu-

6. Anforderungen an das digitale Objekt

- Der transkribierte Text muss in der definitiven Fassung in dem für die Erstellung verwendeten Dateiformat vorliegen.
- Es muss die verwendete Applikation inkl. Versionsangabe und das genutzte Betriebssystem angegeben werden.
- Der transkribierte Text muss zudem als Unicode Text-Datei in der UTF-8-Kodierung vorliegen. Diese Datei enthält die definitive Fassung des Transkripts und soll aus dem Dateiformat, das für die Erstellung verwendet wurde, heraus generiert werden.
- Alternativ zur Unicode-Text-Fassung kann die Transkription auch als XML-Datei gemäss dem TEI- oder ALTO-Standard vorliegen. Auch in diesem Fall ist Unicode UTF-8 für die Zeichenkodierung zu verwenden.
- Der transkribierte Text kann zudem in als PDF-Datei, vorzugsweise PDF/A, vorliegen.
- Fussnoten und Anmerkungen müssen in der Unicode-Text-Fassung des Transkripts wiedergegeben werden können. Nach der Umwandlung ist dies zu überprüfen.
- Die Unicode-Text-Fassung der Transkription darf keine Elemente aus der Kopf- und Fusszeile von gängigen Officeanwendungen enthalten. Dies ist nach der Umwandlung zu überprüfen und ggf. durch Nachbearbeitung zu erreichen.
- Wird die Transkription zeilenidentisch erstellt, muss bei der Umwandlung in die Unicode-Text-Fassung der Zeilenumbruch beibehalten werden. Es soll in diesem Fall auch angegeben werden, wie der Zeilenumbruch definiert ist: Mit Wagenrücklauf/carriage return (CR) oder Zeilenvorschub/line feed (LF) oder mit beiden und wenn letzteres, in welcher Reihenfolge.
- Wenn

7. Anforderungen an den digitalen Text

- Die Überführung des handschriftlichen in einen digitalen Text ist möglichst genau vorzunehmen. Es soll eine weitgehend diplomatische Umschrift erfolgen, d. h. Normalisierung und Modernisierung des Textes sollen möglichst gering ausfallen.
- Abweichungen von den nachstehenden Transkriptionsregeln sind zu benennen.
- Das verwendete Regelwerk ist über den gesamten Text einheitlich anzuwenden.
- Die verwendete Vorlage ist eindeutig zu bezeichnen; insbesondere ist die Archivsignatur, ggf. die URL des Digitalisats anzugeben.
- Bildet ein Digitalisat des Staatsarchivs die Vorlage für die Edition resp. Transkription, ist dies anzumerken.
- Bei der Seitenzählung ist neben der Seitenzählung (inkl. ergänzte Seiten durch den Herausgeber) auch die Scannummer anzugeben. Dabei steht die Scannummer nach der Seitenzählung in eckigen Klammern [scan_xxx].

8. Anforderungen an die Transkription

8.1. Textgestaltung

- Die Seitenzählung ist beizubehalten; fehlende Paginierung ist zu ergänzen und die ergänzten Seitenzahlen in eckige Klammern [] zu setzen.

mentation, and are published under an open-source license." (<http://www.tei-c.org/Guidelines/>) [besucht am 11.4.2012]. ALTO ist die Abkürzung für "Analyzed Layout and Text Object". ALTO ist ein "XML Schema that details technical metadata for describing the layout and content of physical text resources, such as pages of a book or a newspaper. It most commonly serves as an extension schema used within the [Metadata Encoding and Transmission Schema \(METS\) administrative metadata section](#). However, ALTO instances can also exist as a standalone document used independently of METS" (<http://www.loc.gov/standards/alto/>) [besucht am 11.4.2012].

- Seitenzahlen werden zu Beginn einer neuen Seite der Vorlage auf der ersten Zeile platziert.
- Der Seitenwechsel der Vorlage wird durch doppelten Schrägstrich // gefolgt von einer Zeilenschaltung markiert.
- Der Zeilenwechsel der Vorlage wird durch einfachen Schrägstrich / markiert.
- Die Markierungen für Seite- und Zeilenwechsel werden auch gesetzt, wenn die Transkription seiten- und zeilenidentisch erfolgt.
- Titelblätter werden zeilenidentisch wiedergegeben.
- Wenn eine Zeilenzählung verwendet wird, soll diese durch eckige Klammern als Zusatz des Herausgebers gekennzeichnet sein und als eindeutiges Kennzeichen als erstes Zeichen innerhalb der Klammer ein kleines z enthalten.

8.2. Buchstaben, Zahlen und Zeichen

- Grundregel ist die buchstabengetreue Wiedergabe.
- Verschreibungen und Fehler werden wiedergegeben und im Kommentar vermerkt.
- Schreibvarianten einzelner Buchstaben werden nicht wiedergegeben.
- Diakritische Zeichen und übergeschriebene Buchstaben werden buchstabengetreu wiedergegeben. Alternativ können diese wie folgt normalisiert werden:
 - Übergeschriebene Buchstaben werden nachgestellt und durch vorangestelltes # gekennzeichnet.
 - Übergeschriebene zweiteilige Zeichen werden als " wiedergegeben,
 - Übergeschriebene einteilige Zeichen werden buchstabengetreu wiedergegeben.
 - Nicht eindeutig identifizierbare
- Ligaturen können wiedergegeben oder aufgelöst werden. Das verwendete Verfahren ist anzugeben.
- Kurzes oder rundes und langes oder Schaft- s werden nicht differenziert.
- ß wird wiedergegeben und von sz unterschieden. Im Zweifelsfall wird sz verwendet.
- Die Buchstabenkombinationen cz und tz werden gemäss Vorlage wiedergegeben. Im Zweifelsfall wird tz verwendet.
- Die Buchstabenkombinationen dz, dc, wz werden gemäss Vorlage wiedergegeben oder aufgelöst. Das verwendete Verfahren ist anzugeben.
- Zwei Punkte (ÿ) oder ein Strich über y werden weggelassen; der Buchstabe wird als y wiedergegeben.
- u und v werden gemäss der Vorlage wiedergegeben.
- i und j werden gemäss der Vorlage wiedergegeben.
- Der Geminationsstrich bei n und m wird in nn und mm aufgelöst.
- Distinktionszeichen über u, die eine Verwechslung mit n verhindern sollen, werden nicht wiedergegeben.
- römische und arabische Zahlen werden gemäss der Vorlage wiedergegeben. Römische Zahlen werden dabei stets klein geschrieben.
- Gedankenstrich (--), Trennungsstrich (-), Bis-Strich (--) entsprechen dem heutigen Zeichenbestand.
- Ein doppelter Abkürzungspunkt ist durch einen einfachen Punkt wiederzugeben.
- Anführungs- und Schlusszeichen sind durch gerade Anführungs- und Schlusszeichen wiederzugeben. Typografische Anführungs- und Schlusszeichen dürfen nicht verwendet werden.

8.3. Interpunktion

- Die Interpunktion der Vorlage - auch im Wortinneren und bei Ordinal- und Kardinalzahlen - ist unverändert zu übernehmen.
- Die Virgel ist durch Komma zu ersetzen.

8.4. Wortabstand

- Ein vorhandener Wortabstand wird durch ein Leerzeichen wiedergegeben.
- die Variabilität des Wortabstands in der Handschrift ist bei der Transkription zu berücksichtigen.

8.5. Gross- und Kleinschreibung

- Es wird die Schreibung der Vorlage übernommen.
- Wenn die Buchstabenformen nicht eindeutig zu unterscheiden sind, ist Kleinschreibung zu verwenden.
- Grossschreibung erfolgt abweichend immer am Satzanfang und bei Eigennamen. Unsicherheiten sind anzumerken.

8.6. Zusammen- und Getrennschreibung

- Die Schreibung der Vorlage wird übernommen. In Zweifelsfällen kann die heutige Schreibweise verwendet werden; sie sind anzumerken.

8.7. Hervorhebungen

- Verlängerte Buchstaben, Unterstreichungen, Sperrungen, Farben werden nicht dargestellt. sie können in Anmerkungen nachgewiesen werden.
- Bei einer Erfassung mittels TEI oder ALTO sind Hervorhebungen über Auszeichnungselemente wiederzugeben.

8.8. lateinische Schrift (antiqua)

- in Texten mit deutscher Kurrentschrift als Grundschrift werden Inhalte in lateinischer Schrift *kursiv* wiedergegeben.
- Bei der Umwandlung in Unicode-Text-Datei kann die Auszeichnung der unter a. genannten Inhalte als kursive resp. Antiqua über die Einfügung der folgenden Auszeichnungselemente erfolgen: <kursiv> zu Beginn der entsprechenden Textpassage und </kursiv> am Ende der entsprechenden Textpassage.
- Bei einer Erfassung mittels TEI oder ALTO sind kursiv wiedergegebene Textpassagen über Auszeichnungselemente wiederzugeben.

8.9. Abkürzungen

- Abkürzungen, die das Textverständnis nicht erschweren, werden wiedergegeben.
- Alle anderen Abkürzungen werden aufgelöst.
- unsichere Auflösungen werden durch eckige Klammern gekennzeichnet, wobei zwecks Volltextsuche das gesamte Wort in eckige Klammern gesetzt wird und nicht nur der aufgelöste Teil.
- Der Abkürzungspunkt entfällt bei der Auflösung.
- konventionelle Kürzungszeichen am Wortende werden stillschweigend aufgelöst.

8.10. Korrekturen Ergänzungen, Streichungen oder Anmerkungen

- Streichungen in der Vorlage werden transkribiert und in geschweifte Klammern { } gesetzt.
- Ergänzungen, auch Korrekturen, in der Vorlage werden in spitze Klammern < > gesetzt.
- Ergänzungen am Rande (Marginalien) werden ebenfalls in spitze Klammern < > gesetzt und in den Text aufgenommen. Mittels Kommentar wird nachgewiesen, dass es sich um eine Marginalie handelt.
- Streichungen und Ergänzungen von späterer Hand sind im Kommentar als solche nachzuweisen.
- Fussnoten und Anmerkungen der Vorlage sind als solche wiederzugeben. Die originale Nummerierung ist wiederzugeben. Sie sind als originale Fussnoten und Anmerkungen zu kennzeichnen, sofern Fussnoten und Anmerkungen durch den Herausgeber vorhanden sind.
- Bearbeitungsspuren auf der Vorlage sind als Ergänzungen oder Streichungen zu behandeln.

8.11. Lücken, Mängel und Leerstellen in der Vorlage

- Lücken in der Vorlage, die auf Textverderbnis zurückzuführen sind, sind mit [xxx] zu kennzeichnen und können in einer Anmerkung genauer beschrieben werden (Rasur, Frassspuren, Brandloch, Tintenfleck).
- Unleserliche Passagen sind mit [unleserlich] zu kennzeichnen und können in einer Anmerkung genauer beschrieben werden.
- Leerstellen und Leerblätter in der Vorlage sind durch drei Punkte ... zu kennzeichnen und der Umfang in einer Anmerkung festzuhalten.

8.12. Zusätze des Herausgebers

- alle Zusätze des Herausgebers sind in eckige Klammern [] zu setzen.
- Unsichere Lesarten von Wörtern können durch nachgestelltes [?] gekennzeichnet und in Anmerkungen genauer beschrieben werden.
- Auslassungen werden durch [...] gekennzeichnet. Der Umfang der Auslassung ist in einer Anmerkung nachzuweisen.
- In der Vorlage fehlende Buchstaben können ergänzt werden. Die ergänzten Buchstaben müssen durch eckige Klammern gekennzeichnet werden.
- Fussnoten und Anmerkungen durch den Herausgeber werden in eckige Klammern gesetzt. Die Fussnotenzeichen sind ebenfalls in eckige Klammern zu setzen.